PROCEEDINGS OF

The 61st SIMS Conference
on Simulation and Modelling
SIMS 2020

Editors: Esko Juuso, Bernt Lie, Erik Dahlquist, and Jari Ruuska

Organized by University of Oulu
Finnish Simulation Forum (FinSim)
and
Finnish Society of Automation

Proceedings of
The 61st SIMS Conference on Simulation and Modelling
SIMS 2020

Virtual Conference, Finland, 22-24 September 2020

Editors:

Esko Juuso, Bernt Lie, Erik Dahlquist, and Jari Ruuska

Organized by:

University of Oulu

Finnish Simulation Forum (FinSim)

and

Finnish Society of Automation

in cooperation with:
Scandinavian Simulation Society (SIMS)

Preface

The 61st SIMS conference on Simulation and Modelling (SIMS 2020) was organized as a virtual conference. Originally, this conference was planned to be the first SIMS EUROSIM conference and organized in Oulu, Finland. The COVID-19 pandemic presented tremendous challenges for the global research community and for the entire world. The organizers were first postponing the deadlines and keeping the plan of organizing the conference in person. Since the pandemic was continuing strongly, the plans were changed: the SIMS 2020 was decided to organize as a virtual conference in September 2020 and the first SIMS EUROSIM conference in September 2021.

The Scandinavian Simulation Society consists of members from five Nordic countries: Denmark, Finland, Norway, Sweden, and Iceland. The goal of SIMS is to further develop the science and practice of modelling and simulation in all application areas and to be a forum for information interchange between professionals and non-professionals in the Nordic countries. SIMS is a member society of The Federation of European Simulation Societies (EUROSIM) was set up in 1989. The purpose of EUROSIM is to provide a European forum for regional and national simulation societies to promote the advancement of modelling and simulation in industry, research and development. EUROSIM consists of 17 European Simulation Societies. The Scandinavian simulation society (SIMS) had Board and Annual Meetings during the conference. The conference program consisted of four keynote presentations, a SIMS history review, 70 regular presentations and a panel discussion. The proceedings include 68 full papers. The keynotes and the history review are included as abstracts. The call for papers resulted in 81 submissions prepared by 181 authors from eleven countries. The reviews of all submissions were done by four chairs, eleven IPC members and 37 international reviewers. Full papers were selected on the grounds of academic merit and relevance to the conference theme. Each submission had 2-4 reviews and the acceptance rate was 84% for the full papers published in the proceedings.

The SIMS 61 conference covered broad aspects of simulation, modeling and optimization in engineering applications, including many papers on energy, industry, circular economy, automation and methodologies. Energy papers focus on buildings, district heating, hydro power and heat production. Industry includes papers on steel industry, material processing, granulation, oil and gas industry and transportation. Chircular economy related papers study anaerobic digestion, pyrolysis, condensation and microbial electrosynthesis. The part of CO2 capture and use include CO2 injection, heat exchangers in CO2 capture and CO2 absorption. Automation related papers focus on monitoring and control. Digital twins are studied in industrial processes, fault detection and ERP and MES. Fluid dynamic part include CFD and CPFD models, computational studies of fluidized beds, a gasification reactor and subcooled boiling heat transfer. Statistical and intelligent methodologies papers use the Bayesian approach, partial least squares, intelligent methodologies and agent-based simulations.

Panel discussions were organised on future challenges and possibilities for simulation. The discussion focused on five areas: simulation, energy systems, big data, environment and simulation toolboxes. The virtual conference did not include technical tours. Industrial and environmental applications, development of modelling and simulation tools and strong support for PhD students continue for stimulating process development model-based automation.

We would like to express our sincere thanks to the keynote speakers, authors, session chairs, members of the program committee and additional reviewers who made this conference such an outstanding success. Finally, we hope that you will find the proceedings to be a valuable resource in your professional, research, and educational activities whether you are a student, academic, researcher, or a practicing professional.

Esko Juuso, Bernt Lie, Erik Dahlquist, and Jari Ruuska

Table of Contents

Conferences location

The conference was organized as a virtual event.

Opening, 22 September 2020

Opening of the 60th International Conference of Scandinavian Simulation Society (SIMS): SIMS 60 years

- SIMS President, Prof. Bernt Lie, University of South-Eastern Norway, Norway
- Adj. prof. Esko Juuso, Conference Chair, University of Oulu, Finland

Keynote presentations

Overview and Outlook for the OpenModelica Environment and its Use for Cyber-physical System Development
Peter Fritzson, Department of Computer and Information Science, Linköping University, Sweden

Real-time optimization and control with inaccurate model
Prof. Sebastian Engell, Process Dynamics and Operations Group, Biochemical and Chemical Engineering Department TU Dortmund, Germany

Digital Twins utilization throughout the Life Cycle of Industrial Processes
Dr. Tuula Ruokonen, Digital Services Solutions in Valmet Technologies Oy, Finland

SIMS 60+

Scandinavian Simulation Society 60+ ready for future challenges
Esko Juuso, University of Oulu, Finland,

Conference topics

The Proceedings includes 68 articles in five tracks including 15 topics:

Tracks	Topics	Pages
Energy	Energy in buildings	1 - 38
	District heating	39 - 84
	Heat energy	85 - 115
	Hydro power	116 - 138
Industry	Steel industry and material processing	139 - 164
	Granulation process	165 - 194
	Process development	195 - 228
	Oil and gas industry	229 - 256
	Transportation	257 - 286
Circular economy	Separation and synthesis	287 - 310
	CO_2 capture and use	311 - 337
Automation	Monitoring and control	338 - 364
	Digital twins	365 - 384
Methodologies	Fluid dynamics	385 - 442
	Statistical and intelligent methodologies	443 - 485

Panel discussion on Future challenges and possibilities for simulation, 24 September 2020

Chair: Adj. prof. Jari Ruuska, University of Oulu, Finland

Panelists:
  Prof. Peter Fritzson, Linköping University, Sweden
  Prof. Sebastian Engell, TU Dortmund, Germany
  Dr. Tuula Ruokonen, Valmet Technologies Oy, Finland
  Prof. Bernt Lie, University of South-Eastern Norway, Norway
  Senior prof. Erik Dahlquist, Mälardalen University, Sweden
  Adj. prof. Esko Juuso, University of Oulu, Finland

Conference program

Each conference day started with a keynote and continued with 2 – 3 parallel sessions. The Annual SIMS meeting was in the end of the second day. The third day ended with the SIMS 60+ presentation and the panel discussion. More information is available at SIMS website (https://www.scansims.org/).

Conference General Chair

Adjuct prof. Esko Juuso, University of Oulu, Finland

International Program Committee

Prof. Bernt Lie, University of South-Eastern Norway, Norway, Chair

Adj. prof. Esko Juuso, University of Oulu, Finland, Co-Chair

Prof. Erik Dahlquist, Malardalen University, Sweden, Co-Chair

Adj. prof. Jari Ruuska, University of Oulu, Finland, Co-Chair

Dr. Andreas Körner, Vienna University of Technology, Austria

Prof. Juan Ignacio Latorre-Biel, Public University of Navarre, Spain

Prof. Kauko Leiviskä, University of Oulu, Finland

Prof. Miguel Mujica-Mota, Amsterdam University of Applied Sciences, The Netherlands

Adj. prof. Esa Muurinen, University of Oulu, Finland

Dr. Markku Ohenoja, University of Oulu, Finland

Prof. Kim Sörensen, Alborg University, Denmark

Dr. Satu Tamminen, University of Oulu, Finland

Adj. prof. Kai Zenger, Aalto University, Finland

Prof. Borut Zupančič, University of Ljubljana, Slovenia

Prof. Lars Erik Øi, University of South-Eastern Norway, Norway

National Organizing Committee

Adj. prof. Jari Ruuska, University of Oulu, Finland, Chair

Adj. prof. Esko Juuso, University of Oulu, Finland, Co-Chair

Ms. Anu Randén-Siippainen, Finnish Automation Society, Finland

Mr. Marko Vuorio, Finnish Automation Society, Finland

International Reviewers

Title	Givenname	Surname	Affiliation	Country
Assoc. prof.	Saleh	Alaliyat	Norwegian University of Science and Technology	Norway
Assist. prof.	Ioanna	Aslanidou	Mälardalen University	Sweden
Dr.	Christian	Berg	Kelda dynamics	Norway
Dr.	Ole Magnus	Brastein	University of South-Eastern Norway	Norway
M.Sc.	Khim	Chhantyal	National Oilwell Varco	Norway
Assoc. prof.	Alfredo	Carella	Oslo Metropolitan University	Norway
Assoc. prof.	Ali	Ghaderi	University of South-Eastern Norway	Norway
Assoc. prof.	Per Morten	Hansen	University of South-Eastern Norway	Norway
Prof.	Finn Aakre	Haugen	University of South-Eastern Norway	Norway
M.Sc.	Petri	Hietaharju	University of Oulu	Finland
Prof.	Yrjö	Hiltunen	University of Eastern Finland	Finland
M.Sc.	Antti	Koistinen	University of Oulu	Finland
Prof.	Hailong	Li	Mälardalen University	Sweden
Assoc. prof.	Joachim	Lundberg	University of South-Eastern Norway	Norway
M.Sc.	Heidi	Marais	Mälardalen University	Sweden
Prof.	Britt	Moldestad	University of South-Eastern Norway	Norway
Docent	Esa	Muurinen	University of Oulu	Finland
M.Sc.	Riku-Pekka	Nikula	University of Oulu	Finland
Assoc. Prof.	Lars O.	Nord	Norwegian University of Science and Technology	Norway
Dr.	Eva	Nordlander	Mälardalen University	Sweden
Assoc. prof.	Ottar	Osen	Norwegian University of Science and Technology	Norway
M.Sc.	Madhusudhan	Pandey	University of South-Eastern Norway	Norway
Dr.	Aarne	Pohjonen	University of Oulu	Finland
Prof.	Mika	Ruusunen	University of Oulu	Finland
M.Sc.	Outi	Ruusunen	University of Oulu	Finland
Dr.	Jouni	Savolainen	Sementum Ltd.	Finland
Dr.	Roshan	Sharma	University of South-Eastern Norway	Norway
Prof.	Nils-Olav	Skeie	University of South-Eastern Norway	Norway
M.Sc.	Aleksander	Skrede	Norwegian University of Science and Technology	Norway
Dr.	Aki	Sorsa	University of Oulu	Finland
Prof.	Lars-Andre	Tokheim	University of South-Eastern Norway	Norway
Dr.	Jani	Tomperi	University of Oulu	Finland
Prof.	Knut	Vaagsaether	University of South-Eastern Norway	Norway
Dr.	Ari	Vuokila	University of Oulu	Finland
Prof.	Anis	Yazidi	Oslo Metropolitan University	Norway
Dr.	Peter	Ylen	VTT Technical Research Centre of Finland	Finland
Dr.	Nathan	Zimmerman	Mälardalen University	Sweden

Overview and Outlook for the OpenModelica Environment and its Use for Cyber-physical System Development

Peter Fritzson

Department of Computer and Information Science
Linköping University, Linköping, Sweden

Abstract

The industry is currently seeing a rapid development of cyber-physical system products containing integrated software, hardware, and communication components. The increasing system complexity in the automotive and aerospace industries are some examples. The systems, that are developed, have increasing demands of dependability and usability. Moreover, lead time and cost efficiency continue to be essential for industry competitiveness. The extensive use of modeling and simulation - Model-Based Systems Engineering tools - throughout the value chain and system life-cycle is one of the most important ways to effectively target these challenges. Simultaneously there is an increased interest in open source tools that allow more control of tool features and support, and increased cooperation and shared access to knowledge and innovations between organizations.

Modelica is a modern, strongly typed, declarative, equation-based, and object-oriented (EOO) language for model-based systems engineering including modeling and simulation of complex cyber-physical systems Major features are: ease of use, visual design of models with a combination of lego-like predefined model building blocks, ability to define model libraries with reusable components, support for modeling and simulation of complex applications involving parts from several application domains, and many more useful facilities. The Modelica language is ideally suited for cyber-physical modeling tasks since it allows integrated modelling of discrete-time (embedded control software) and continuous-time (process dynamics, often for physical hardware). Modelica 3.3 extended the language with clocked synchronous constructs, which are especially well suited to model and integrate physical and digital hardware with model-based software.

This talk gives an overview and outlook of the OpenModelica environment – the most complete Modelica open-source tool for modeling, engineering, simulation, and development of systems applications (www.openmodelica.org), and its usage for cyber-physical system development. Special features are MetaModeling for efficient model transformations, debugging support for equation-based models, support (via OMSimulator) for the Functional Mockup Interface for general tool integration and model export/import between tools, model-based optimization, as well as generation of parallel code for multi-core architectures.

Moreover, also mentioned is recent work to make an OpenModelica based tool chain for developing digital controller software for embedded systems, and in generating embedded controller code for very small target platforms like Arduino Boards with down to 2 kbyte memory. This work is extended in the ongoing EMPHYSIS project where the FMI standard is extended into the eFMI standard for embedded systems.
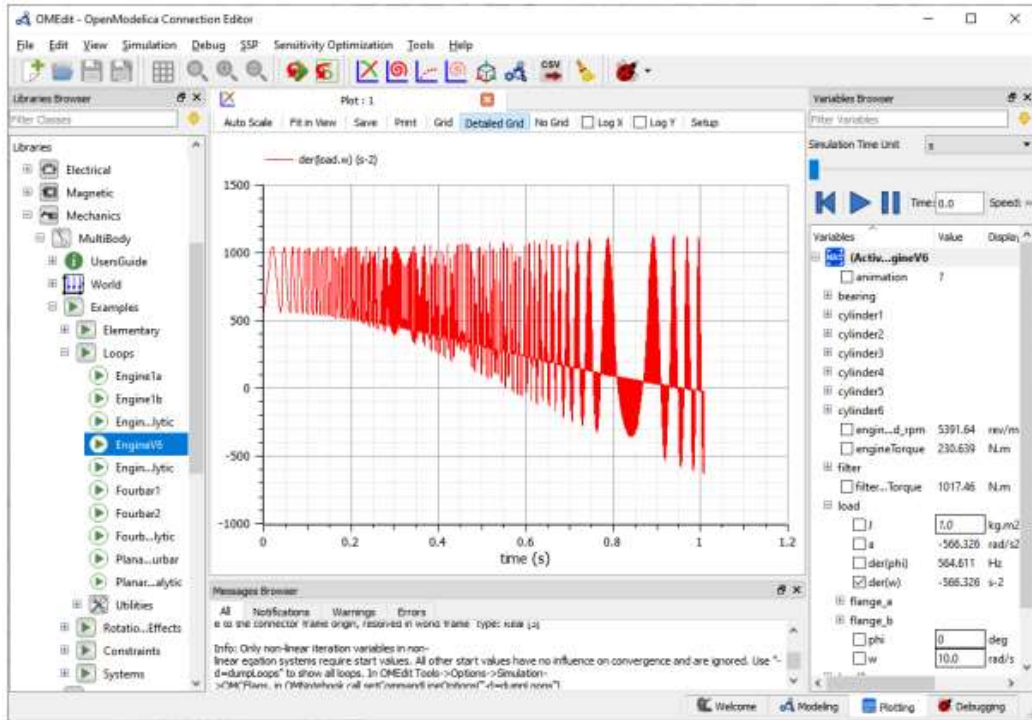
Figure 1. OpenModelica simulation of the V6Engine model with 11000 equations. Plotting simulation results using OMEdit. Left: Model browser. Right: Plot variable browser. Bottom: message browser window.
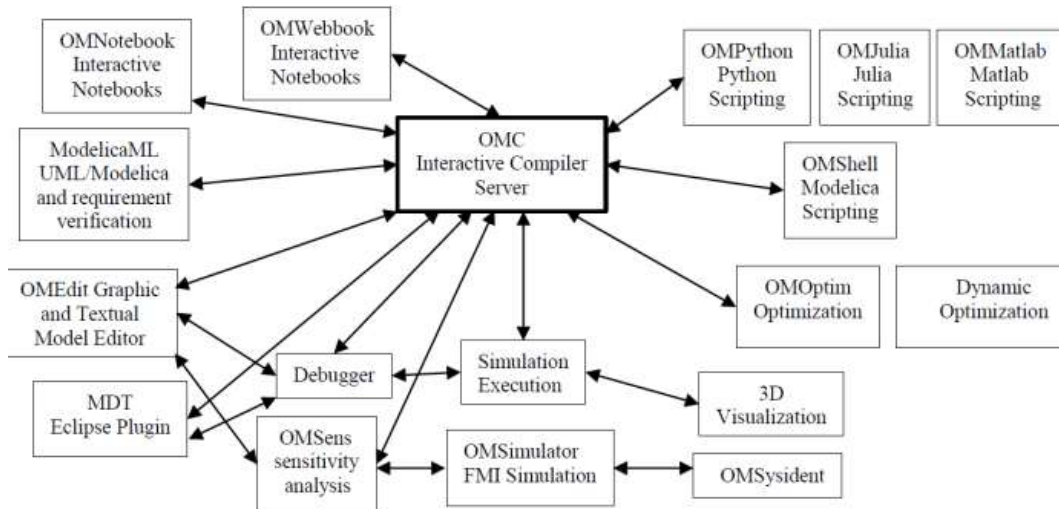

Figure 2. The architecture of the OpenModelica environment. Arrows denote data and control flow.

Biography

Peter Fritzson is the professor and research director of the Programming Environment Laboratory, at Linköping University (LiU). He is also the vice director of the Open Source Modelica Consortium, the vice director of the MODPROD center for model-based product development, (previously the director of both) organizations he took initiative to establish. During 1999-2007, he served as chairman of the Scandinavian Simulation Society, and secretary of the European simulation organization, EuroSim. During 2000-2020, he was vice Chairman of the Modelica Association. Prof. Fritzson's current research interests are in software technology, especially programming languages, tools and environments; parallel and multi-core computing; compilers and compiler generators, high level specification and modeling languages with special emphasis on tools for object-oriented modeling and simulation where he is one of the main contributors and founders of the Modelica language.

He received PhD degree in Computer Science at LiU in 1984. The thesis title was "Towards a Distributed Programming Environment Based on Incremental Compilation". In 1985-86, he was the project leader of the subproject on incremental compilation at Sun MicroSystems, Mountain View, California. In 1989, he became associate professor and director of the Programming Environment Laboratory (PELAB) in Department of Computer and Information Science at LiU. In 1992, he qualified as docent at LiU. From 1995 to 2019 he has been a full professor at LiU and continued from 2020 as a full professor emeritus at LiU. The research director position of PELAB he has had since 1995. Prof. Fritzson has many positions in scientific organizations: the vice chairman of the Modelica Association since 1999, the director of the Open Source Modelica Consortium since 2007, and the research leader at RISE, SICS East, St Anna Research Institute since 2007.

Professor Fritzson has authored or co-authored 319 technical publications, including 21 books/proceedings, 1 book draft, 45 journal papers, and 251 papers in conference proceedings or external book chapters, of which 249 are refereed, 5 are invited, and 2 are nonrefereed. Google scholar link: https://scholar.google.com/citations?authuser=1&user=KXUHll4AAAAJ

Selected books:

Peter Fritzson. *Principles of Object Oriented Modeling and Simulation with Modelica 2.1*. 940 pp, Wiley-IEEE Press, 2004.

Peter Fritzson. *Introduction to Modeling and Simulation of Technical and Physical Systems with Modelica*. 232 pages, ISBN: 978-1-1180-1068-6, Wiley-IEEE Press, September, 2011

Peter Fritzson. *Principles of Object Oriented Modeling and Simulation with Modelica 3.3: A Cyber-Physical Approach*. 1250 pages. ISBN 9781-118-859124, Wiley IEEE Press, 2014.

Real-time optimization and control with inaccurate models

Prof. Sebastian Engell

Lehrstuhl für Systemdynamik und Prozessführung / Process Dynamics and Operations Group
Fakultät Bio- und Chemieingenieurwesen / Biochemical and Chemical Engineering Department
TU Dortmund, Dortmund, Germany

Abstract

Modelling for simulation, modelling for optimization, and modelling for control follow the same principles, but have to meet different requirements. Simulation models should represent the behavior of the system under consideration for a predefined set of test cases faithfully, with the accuracy usually being measured in the time domain, i.e. by looking at the differences of the stationary values or trajectories of some key variables. Modelling for optimization is different in that the goal is that the computed optimum of the model and the optimum of the real system should match, and the model should support efficient global numerical optimization. This implies that the model must be accurate around the global optimum, but over the full range of the variables, medium accuracy is sufficient. The problem for modelling of course is that without the knowledge of the region of the global optimum, it is not possible to build such a model, and hence modelling and optimization should be interleaved.

When using models for control, the relationship between model accuracy and control performance is even more intricate than for optimization. One can control a plant satisfactorily using a coarse model (in the simplest case, the sign of the gain is sufficient), while model errors can also lead to poor performance and instability. For linear time-invariant control loops, the classical robust control theory from the 1980s tells us that a good model for control is a model that describes the dynamic behavior accurately near the gain-crossover frequency, and qualitatively correctly for lower frequencies. So what a good model is for the purpose of control depends on how fast we want to control the system.

For high-performance control, techniques based on online optimization, i.e. model-predictive control (MPC) have become the dominant technology in the last decades, due to their ability to handle constraints, nonlinear systems, and economically motivated cost functions (Engell, 2007). In contrast to the classical theory of robust control, robustness to model errors for such control strategies is difficult to analyze and, similar to stability, is usually handled using a constructive approach, i.e. by building controllers that have certain robustness properties for a given description of the uncertainty of the model. Min-max robust MPC, in which the performance is optimized for the worst case model, is the best known representative of this approach. This however comes at the price of a high conservatism.

To build good models is a costly endeavor. Therefore, both in modelling for optimization and in modelling for control, one is interested in techniques that provide good performance without huge modelling efforts. In the presentation, we discuss two recent approaches to reducing the negative effects of model errors in optimization and control. For real-time optimization, we outline the so-called modifier adaptation approach, which adds a data-based local model to a global model and updates it iteratively to ensure convergence to the true optimum of the real plant (Gao and Engell 2005, Gao, Wenzel, and Engell, 2016). For control, the multi-stage MPC approach is discussed in which the future information on the realization of the model uncertainty is included in the optimization that is performed at the current time step to reduce the conservatism of the controller (Lucia, Finkler, and Engell, 2013).

References

S. Engell. Feedback control for optimal process operation. *Journal of Process Control* 17(3):203-219, 2007. doi: 10.1016/j.jprocont.2006.10.011.

W. Gao and S. Engell. Iterative set-point optimization of batch chromatography. *Computers and Chemical Engineering*, 29:1401-1409, 2005.

W. Gao, S. Wenzel, and S. Engell. A reliable modifier-adaptation strategy for real-time optimization. *Computers and Chemical Engineering*, 91:318-328, 2016. doi: 10.1016/j.compchemeng.2016.03.019.

S. Lucia, T. Finkler, ans S. Engell. Multi-stage nonlinear model predictive control applied to a semi-batch polymerization reactor under uncertainty. *Journal of Process Control*, 23:1306-1319, 2013. doi: 10.1016/j.jprocont.2013.08.008.

Biography

Prof. Engell received a Dipl.-Ing degree in Electrical Engineering from Ruhr-Universität Bochum, Germany in 1978 and the Dr.-Ing. Degree and the venia legendi in Automatic Control from Universität Duisburg in 1981 and 1987. 1984/1985 he spent a year as a PostDoc at McGill University, Montréal, Canada. 1986-1990 he was the head of an R&D group at the Fraunhofer Institut IITB in Karlsruhe, Germany. 1990 he was appointed to his present position as a Full Professor of Process Dynamics and Operations at the Department of Chemical Engineering at TU Dortmund. 2008 he was a Distinguished Visiting Professor at Carnegie Mellon University, Pittsburgh, USA. He was Department Chairman 1996-1999 and 2012-2014 and Vice-Rector for Research of TU Dortmund 2002-2006. He is currently a member of the Research Council of the Alliance of the Universities in the Ruhr Region, UA Ruhr.

Prof. Engell received an IFAC Journal of Process Control Best Paper Award, and is a co-author of the 2014 and 2016 Best Papers in Computers and Chemical Engineering. He received the Best Paper Award of the IEEE Congress on Evolutionary Computation 2010 with Thomas Tometzki on risk-conscious planning and the PSE Model-based Innovation Prize with Ala Eldin Bouaswaig. He gave the Bayer Lecture in Process Systems Engineering at Carnegie Mellon University in 2008 and the Roger Sargent Lecture at Imperial College, London, in 2012. He has published more than 120 Papers in scientific journals, more than 40 papers in edited volumes and more than 300 conference papers with peer review and full papers in proceedings. His Scopus paper count is 530 with 4900 citations. He graduated more than 70 PhD students at TU Dortmund. In 2012, he was awarded a European Advanced Investigator Grant for the Project MOBOCON – Model-based Optimizing Control – From a Vision to Industrial Reality.

Prof. Engell is a Fellow of IFAC, the International Federation of Automatic Control since 2006 and has led the IFAC Fellow Selection Committee 2012-2014 He served as President of EUCA, the European Control Association and is a member of the selection committee for the European Control Award.

Prof. Engell has led several European Projects in the FP6, FP7 and Horizon 2020 Frameworks: Multiform (ICT), DYMASOS (ICT), CPSoS (Support Action, ICT), MORE (NMP) and CoPro (SPIRE, ongoing). The CPSoS project developed a roadmap for Cyber-physical Systems of Systems in Europe. He was involved in the Marie Curie Reseach Training Networks oCPS and PRONTO and currently is a partner in the EU-India project LOTUS on monitoring water quality and managing water systems.

His research areas are in the domains of model-based optimizing control, real-time optimization, and scheduling. In his research, the aspect of uncertainty about the behavior of the system that is controlled or optimized has always been in the focus.

Digital Twins utilization throughout the Life Cycle of Industrial Processes

Tuula Ruokonen, Ph.D (Eng), MBA

Director, Digital Services Solutions, in Valmet Technologies Oy;
Board Member in Automation Foundation in Finland sr.

Abstract

In this presentation, Digital Twins for industrial processes are considered from their historical and future point of view. What is actually a Digital Twin – is their only one or several for different purposes? What enables the development of Digital Twins just now? Which benefits and challenges are there in their development and implementation?

There are many definitions for Digital Twins. Most state that a Digital Twin is a virtual representation of a physical product or process, used to understand and predict the physical counterpart's performance characteristics. Digital Twins are used throughout the product life cycle to simulate, predict, and optimize the product and production system before investing in physical prototypes and assets.

Already 30 years ago, such systems and simulators were developed and utilized which for sure would nowadays be called Digital Twins, concrete examples including Computer Aided Design, Process Modelling and Dynamic Simulators, Advanced Process Control, Condition Monitoring, Expert and Knowledge-Based Systems, and even Remote Expert Services.

Digital Twins are presently at the top in their hype curve, and their enabling technologies develop strongly and rapidly. We are facing partly evolution, partly revolution in their development. For example increased computing power enables real time analytics, cloud-based computing enables flexible calculation capacity, mobile technology enables mobile and remote applications, wireless sensors enable additional measurements, and Artificial Intelligence and Machine Learning tools enable advance analytics. Furthermore important is the connection to Internet of Things and Industrial Internet applications development.

Potential to utilize Digital Twins in industrial processes and equipment is wide, them forming ideally a digital thread throughout the whole life cycle. The goal is efficient information management, its utilization and updating in all phases of the life cycle: product development, production planning, sales, project implementation, operations optimization, personnel training, process operation and maintenance.

Why is the utilization of Digital Twins still so difficult or even impossible? Challenges are created by separate functional processes and IT systems, and especially by organizational silos and suboptimization of goals, in different phases of the life cycle. Open questions exist still related to common data models and standards, and model updates. Own challenges come from the data ownership and principles of sharing data between related actors, equipment manufacturers, end users and service providers, related to design data and operation-time data management.

Looking forward that these challenges and open questions will be solved and the vision of up-to-date Digital Twins, utilized in the whole life-cycle, comes true and enables the performance optimization of processes and equipment in the future autonomous mills and plants.

Biography

Dr. Tuula Ruokonen is responsible of development of new value adding applications and service solutions in Valmet, based on remote services, industrial internet and advanced analytics. She is inspired by new business opportunities and applying new technological potentials. Special focus is in testing and taking into use new technologies, such as virtual and augmented reality, and applying AI and machine learning methods. Applications are developed in tight customer co-operation based on design thinking and the Lean Start-up method.

Dr. Ruokonen has more than 25 years of practical business management experience, from technology and product development to line management responsibilities. Especially she has wide experience in industrial service and outsourcing business and its development. She has a Ph.D. degree in Electrical Engineering (Control Engineering), Florida Atlantic University, USA, 1989. Thesis was on "Nonlinear Filtering for Failure Detection in Dynamic Systems". She has M.Sc, and Lic.Tech. degrees in Electrical Engineering (Control Engineering) from the Helsinki University of Technology. The Lic.Tech. thesis (1987) was on "Model-Based Failure Diagnosis - Structure of the System and Methods Based on Characteristic Curves". M.Sc. in Electrical Engineering (major in Control Engineering and Automation, minor in Measurements and Information Technology) was received from the Helsinki University of Technology in 1985. The M.Sc. thesis (1985) was on "Failure Diagnosis and Condition Monitoring in Power Plants. Methods and Applications, a Peat Fuelled Plant as an Example". In 1993, she received MBA degree from the Helsinki University of Technology (Aalto University).

Dr. Ruokonen has received Finnish Society of Automation bronze medal of merit 1996 and silver medal of merit 2000. She is active in automation societies, including Automation Foundation, a board member 1996-98, 2019-, Automaatioväylä magazine, a board member 1991-92, and Finnish Society of Automation, a board member 1990-92. She has been active in The Association of Electrical Engineers in Finland, a board member 1987, 1994-96, and Sähköklubi, a board member 1984, 1989-91, 2017-2019. Already during her post graduate, she was active in the IEEE student branch in Florida Atlantic University, a board member 1987-88.

Valmet is the leading global developer and supplier of technologies, automation and services for the pulp, paper and energy industries, having strong growth strategy and the aim to utilize efficiently new digitalization enablers for customer value-add and internal efficiency. Valmet's net sales in 2019 were approximately EUR 3.5 billion and it employed about 13.600 professionals around the world.

Scandinavian Simulation Society 60+
ready for future challenges

Adj. prof. Esko K. Juuso

Control Engineering, Environmental and Chemical Engineering
Faculty of Technology
University of Oulu, Finland

Abstract

Scandinavian Simulation Society (SIMS) started over 60 years ago as a society for the analog simulation: Scandinaviska Analogmaskinsällskabet (SAMS) was founded in 1959 in Västerås. The current name was taken in 1968 when SIMS moved successfully to the digital simulation. SIMS is currently a society of societies which operates in five countries Denmark, Finland, Iceland, Norway and Sweden. SAMS started in Sweden, Denmark and Norway. Finland joined in the 60s and in 1972 SIMS conference was organized the first time in Finland. In 2012, the SIMS conference went to Iceland. SIMS is the eldest active simulation society in the world. After 60 annual conferences, the 61st conference is the first virtual conference.

Applications have continuously been important parts of the conferences. Steel industry, flight simulators and atomic energy were active already in the beginning. Industry, Energy and Environment are important areas of SIMS. Applications in the energy field have extended from power plants to sustainable energy: solar, wind and geothermal, especially in Iceland. Processes, including the forest, steel and chemical industry, as well as oil & gas production have kept an important role. Increasing interests have been seen in water and wastewater treatment, biogas production and bioprocesses. The annual conferences circulate sequentially in the SIMS countries and the topics adapt in the local interests. SIMS has stimulated automation in these versatile fields, also through representative organizations like Finnish Automation, Automation region in Sweden and Norwegian Automation.

Numerical methods and combined differential and algebraic equations have formed the basis for process models first in Fortran, then Matlab and Simulink. Computational intelligence, AI and various model builders, e.g. gPROMS, became active in the 90s. The full range from PCs to supercomputers is used. The developments of new simulation tools have been important during the years: Apros was introduced in 1986 and Modelica in 1996. As a spinoff, we have the active Modelica Association. Graphics becomes more advanced, open source codes are coming more popular and the integration of methodologies and tools developed for different applications areas extend the application areas.

SIMS provides strong **support for PhD students** by bringing together different methodologies, applications, software tools and people. Efficient simulation tools help in bringing different ideas within the education.

International cooperation has been essential throughout the years. An agreement with International Association for Mathematics and Computers in Simulation (IMACS) dates back 1976 resulting the IMACS World Congress in Oslo 1985. The European Simulation Multiconference was organized in 1991 in Copenhagen. SIMS is an active member society in the Federation of the European Simulation Societies (EUROSIM) founded in 1992. The EUROSIM congress has been organized twice by SIMS: 1998 Helsinki and 2016 Oulu. The SIMS EUROSIM 2021 is starting a new EUROSIM conference series to be held every third year.

The ***future*** can be based on the stimulated use of new simulation tools, a wide scale of applications in industry, energy and environment, all stimulated by the automation and young generation of researchers. This combination, which has kept SIMS going for decades, needs to be adapted in the future challenges.

The ***circular economy*** aims to close the loop to make economy more sustainable and competitive. We have a broad range of technologies related to recycling, renewable energy, information technology, green transportation, electric motors, green chemistry, lighting, grey water, and more. The environment is restored with pollution removal and avoidance. What can we do in practice? Air has been a focus area in industry,

energy and traffic. Water treatment has been developed to remove undesirable chemicals, biological contaminants, suspended solids and gases from contaminated water. In industrial processes, closed water circulation is a goal which is beneficial for the environment. Wastewater treatment is needed for purifying contaminated water before returning it to the nature. Is climate change discussion sufficient? Should we take a wide view on the ecosystem?

The *energy sector* has new challenges and possibilities. Thermal power plants are by far, the most conventional method of generating electric power with reasonably high efficiency. Bioenergy takes an increasing portion of the production? Oil and gas hold a strong position in overall energy usage. Biofuels provide new competing alternatives and the CO_2 capture has taken a high role in research. Are we going to bioeconomy? Is the thermal power a necessity in our energy balance? Sustainable or renewable energy is considered as a future source of energy: water power is well integrated in the energy system; solar and wind are getting more popular; geothermal, wave and tide energy can be locally very important. Electricity is increasingly popular both in solar and wind power. To what level is this sufficient? Where do we use energy? Industry needs high reliable levels. Is the nuclear power a solution? Adaptation is easier in domestic use, but how to do it? Heating and cooling take the highest part. Solar energy can help but needs storage. Geothermal can be used as storage. What is the potential of buildings as storages? Do we need small scale CHP? District heating systems are good solutions to bring the thermal energy to buildings.

In the *industry*, intelligent systems have been developed for integrating data and expertise to develop smart adaptive applications. Recently, big data, cloud computing and data analysis has been presented as a solution for all kinds of problems. Can we take this as a general solution for automation? Wireless solutions are improving fast: 3G, 4G, 5G. But can we transfer signals to clouds and store the data? Is this too much? Where is the expertise? Obviously, local calculations are needed. Are they based on intelligent systems? Transport systems are analyzed as discrete event systems to find bottlenecks and avoid risks. Urban traffic is becoming an important area. Autonomous driving is a hot topic. What is needed to embed this in the urban traffic? Are there analogies with industrial systems? What are the main differences between industrial systems and transport systems? Can we use similar control solutions? What can we learn from other areas? Can we find analogies? What is common? Where do we have differences? What kind of models do we need?

Highly complicated systems with various interactions are at hand. What researchers within SIMS community can do? Do we have tools and methodologies to help in solving these problems? The modelling and simulation is coming increasingly important in various fields.

Biography

Senior Research Fellow Esko Juuso has D.Sc. (Tech.) on Control and Systems Engineering from the Department of Process and Environmental Engineering at the University of Oulu. He is an adjunct professor in Computational Intelligence at the University of Oulu. He also has a M.Sc. (Tech.) degree in Technical Physics (Material Physics) from the University of Oulu. He worked earlier several years as a research engineer and process computer analyst in the metal industry. He has been a team leader and a project manager of several research projects on intelligent systems applications. Adjunct professor Juuso is the Fellow of International Society of Condition Monitoring (ISCM), a member of ISCM Management Committee and the Chair of the ISCM Publication Committee. In the Scandinavian Simulation Society (SIMS), he has been a member of the Board of SIMS since 1996 and the President of SIMS 2007-2013. He is a founding member of the Finnish Simulation Forum (FinSim), 2006, and the International Society for Condition Monitoring (ISCM), 2010. He is the chair of SIMS 2020 and he has chaired SIMS conferences in 2002, 2006 and 2010. Since 2006, he has been the SIMS representative in the Board of the Federation of European Simulation Societies (Eurosim), the secretary of Eurosim 2007-2010 and the President of Eurosim 2013-2016.

Panel discussion:
Future challenges and possibilities for simulation

Chair: Adj. prof. Jari Ruuska, University of Oulu, Finland
Panelists:
> *Prof. Peter Fritzson, Linköping University, Sweden*
> *Prof. Sebastian Engell, TU Dortmund, Germany*
> *Dr. Tuula Ruokonen, Valmet Technologies Oy, Finland*
> *Prof. Bernt Lie, University of South-Eastern Norway, Norway*
> *Senior prof. Erik Dahlquist, Mälardalen University, Sweden*
> *Adj. prof. Esko Juuso, University of Oulu, Finland*

The panel discussion was the last part the conference. The panelist were the keynote presenters, the current and two past presidents of SIMS, including the chair of the conference. The chair of the national organizing committee was the chair of the panel. The discussion focused on five questions: simulation, energy systems, big data, environment and simulation toolboxes. The questions were presented by the panel chair.

1 How would you define term simulation?

Peter: Simulation answers the questions about the model.

Sebastian: nothing to add.

Tuula: There are often several interpretations about the terms. Simulation is often understood only as process simulation but there are also reliability models considering e.g. maintenance, management, and also business-related models.

Bernt: Experiments with models are used to find out answers – model is a wide concept not only including mathematical models but also pilot plants in principle or economic models

Erik: Model building means including new information in the simulation model to see what would happen. It can also be a non-mathematical tool.

Esko: Tool for comparing alternatives – application related in that sense. We can use it to find best alternatives to use. It is different when you are controlling something, designing something, or finding an optimal control point or balance. One should consider "how detailed simulation you need for your job?"

2 Energy systems are needed everywhere but demand of energy changes a lot between seasons and time of the day – How to use simulations to aid in this subject and predict changes?

Peter: You can include all the energy sources in your simulation including the consumers to create a comprehensive model. Historical data about the location's energy usage in conjunction with real-time weather data can be used to generate good predictions.

Sebastian: You have individual systems that are coupled. It is quite a big task for modeling, simulation, and optimization to co-ordinate such individual systems. German airports have saved lots of CO_2 by optimizing airplane approaches to the airport. Philosophically it is more beneficial to have a good aggregate model that you can use for decision making between the models than a huge, detailed model about everything. How to do this in an optimal way is very interesting research question.

Tuula: Distributed energy production and usage; Low level decision making for individual needs. IoT with plenty of measurements – energy usage of all the members and their effect on costs. Big savings are potential in total optimization.

Bernt: Future energy is related to United Nations' goals to provide energy for everybody. A combined system where all the energy sources work better together. Circular economy where energy streams are more integrated: one operator's waste is a raw material for another. This needs simulation tools – short & long -term simulations

Erik: Energy technologies develop in long time perspective, so it is important to consider these long-term changes. Biofuels and batteries will replace oil in transportation systems in future. Great opportunity to use simulation tools here. Boundary determination is more important than very detailed models. Simulation tools can help politicians and decision makers to make right decisions.

Esko: Energy is challenging for control and optimization. Sources that work depending on weather conditions and ones that are not connected to the weather form a diverse system that is combined and have different requirements. Many energy sources that have different fluctuations must be optimized and controlled to work as a complete system. District heating networks are good for balancing different energy sources.

3 Big Data and cloud computing have been presented as a solution for everything. Can we use this as a general solution for automation? - Can we use these solutions reliably considering data security and how to use simulations to aid this?

Peter: Mathematical model equations describing the physical world have developed a lot. Combine these mathematical models to make approximate models for machine learning to get a bigger picture.

Sebastian: Industrial perspective is always skeptical to transfer its data into cloud. There has been work done for MPC over unreliable networks. If there is a problem in considering missing information, it is better to invest in a working network than in sophisticated control solution. There are not many successful online applications of machine learning. Difficult to have sophisticated control that can handle all operating states – you could have good combinations of fundamentals. Models calculating phase equilibrium can take too long to be used in process optimization and control. Approximations using machine learning could be embedded in an overall rigorous model.

Tuula: A great number of applications are already existing for cloud-based solutions. Cyber security is one major risk, a lot of work to be done to cover this topic. Possibility to utilize data from a fleet of the processes can help us understand what the normal operating points is. Training simulators can help to utilize this understanding about the operating points. The future will be more and more in the cloud or on the edge - not only on-site. Big operators like Amazon web services must do their

share in web security.

Bernt: Big data and data driven models are very modern. Physics based models can be built without a real system before building the system to give reasonably accurate prediction. Mechanistic models are typically better outside the operating points than data driven models. Both models have their unique advantages. Combining physical and data driven models – physics-based models for the phenomena you know well and data driven models for the ones you do not know so well.

Erik: We have a project (Future direction of process optimization) focusing on learning systems. There we use infrared measurement for the wood chips lignin content and physics-based model for residual alkali, i.e. how much lignin is there still in fiber. The residual alkali measurement is needed for tuning the model for lignin properties and content. Internal cloud behind the firewalls could be the solution when operators do not trust the providers.

Esko: We need expertise, models, and data driven models for the practical applications. Theoretical models alone are too complex – parameters can be learned from the data. A dangerous thing is to use unbalanced data that gives you a completely wrong answer. Simulation is a good tool when you understand what you are doing. Do not use AI to find answers to questions that experts already know. Finding suitable system size for application is important.

4 Environmental friendliness and sustainability are important but how to use simulation to promote this?

Esko: Simulation can have a big role in this. Combine different ideas and compare different things to select the best solutions – not like now that there are things you should not use and things you should.

Erik: Generate scenarios that give good data for the decision makers.

Bernt: Sufficient energy generation in the world in future needs all energy sources – fossil fuels need to be reduced in future, but it is still needed for now. Complicated systems where we need e.g. to store solar and wind energy needs simulations to see all the consequences.

Tuula: Industrial optimization is important. Optimizing processes is connected with sustainability – stability and energy efficiency. This is cost efficient for the companies and gives

them an advantage, so it is win-win for everyone.

Sebastian: More complex problem that needs to consider the whole life cycle from the raw materials to end of product life which is very problematic. What kind of materials are used for water bottle, how many times this bottle will be used – lots of variables and it is hard to simulate everything.

Peter: Technical simulation gives accurate predictions, but we also need more general simulations. Decision makers (politicians) make lots of decision based on low quality information and intuition – simple, not too detailed models could give more information before making decisions. Not only energy but also materials are important: how can we use alternative materials and lower the material usage.

5   Are there some basic principles that needs to be added into our simulation toolboxes?

Esko: CSC supercomputers can help in calculations, but we need to understand all the parts of the system. We need to do intelligent calculations and understand what we are doing.

Erik: A good formulation of the problem enables e.g. the use of CFD for online applications.

Bernt: A better use of parallelization is needed for networks and graphical processing. Simple automatic way to use this is needed because it is hard if you need to first do everything manually.

Tuula: Calculation tools for online-measurements. Hybrid models are needed for combining data driven models with physical models. Help us in industry to understand what methods to use and where we do not do wrong things.

Sebastian: Quantification of calculation uncertainty is important, and it should be connected to the results.

Peter: Multiparameter sensitivity simulation in Modelica - build uncertainty into tools.

Esko: Uncertainty is unavoidable, therefore it is better to be approximately correct than exactly wrong.

6   Conclusions

The discussion covered well the future challenges and possibilities of simulation. The panel was a highly valuable conclusion for the conference in linking the keynotes and topics of the regular papers with the history and the future of the simulation.

Acknowledgements

Author index

Proceedings of SIMS 2020
Virtual, Finland, 22-24 September 2020

<u>List of Papers</u>

1. Energy

1.1 Energy in buildings

1.2 District heating

1.3 Heat energy

1.4 Hydro power

2. Industry

2.1 Steel industry and material processing

2.2 Granulation process

2.3 Process development

2.4 Oil and gas industry

2.5 Transportation

3. Circular economy

3.1 Separation and synthesis

4. Automation

4.1 Monitoring and control

4.2 Digital twins

5. Methodologies

5.1 Fluid dynamics

Impacts of demand side management programs to domestic hot water heating load profiles in smart buildings

Jari Pulkkinen   Jean-Nicolas Louis

Water, Energy and Environmental Engineering, University of Oulu, Finland,
`jari.pulkkinen@oulu.fi jean-nicolas.louis@oulu.fi`

Abstract

The increasing amount of variable electricity generation has brought world to investigate various flexibility sources to provide power network balancing through demand side management. Therefore, it is important to create new, more thorough models that allow using smart functions to control the various electricity loads. In this paper a model to simulate a fully mixed domestic hot water tank's behavior in 60, 30 and 15 min time resolution, and its control mechanisms were created. The model will be integrated to another smart house model to enable studying more combined smart controls and functions. Additionally, the flexibility of the hot water storage tank was investigated with the help of 4 different heating scenarios, showing its suitability for Demand Side Management, and the operation of the model was confirmed with lower time resolutions.

Keywords: domestic hot water, demand side management, smart building, exploratory scenario

1 Introduction

The increasing electricity generation from variable renewable energy sources creates new requirements for balancing the demand and supply of electricity. To prevent imbalance issues on the network, new kinds of balancing mechanisms are being studied, one of which is to utilize the flexible loads on the demand side to shift electricity consumption according to its supply status. In residential buildings deferrable and thermal loads, like domestic hot water (DHW) storages, are generally considered as good sources for providing flexibility to the network (Lu, 2012). These loads can be used to participate in Demand Side Management (DSM) programs where electricity consumers adjust their electricity consumption to match the status of power network either through price-signals, like Time-of-Use (ToU) or Real-time price (RTP), or by receiving an incentive payment when reacting to a request to reduce electricity consumption (US Department of Energy, 2006). Additionally, combining local electricity generation from e.g. Photovoltaic-panels with flexible electricity loads can help in increasing self-consumption from local electricity generation and provide monetary benefits (Salpakari and Lund, 2016). Therefore, there is a need to identify and study the flexibility in buildings to determine suitable sources and their potentials in participating

in DSM programs, through programs like IEA EBC Annex 67 Energy Flexible Buildings (Jensen et al., 2017).

1.1 Domestic Hot Water

Domestic hot water (DHW) is an important part of having a good living quality and technology to produce and store it are essential in today's world. Hot water is commonly mixed with colder tab water to create thermally comfortable water to be used in shower or as tab water to wash hands. As the consumption of DHW is not constant, it is important to study both their load profiles and the behavior and sizing of DHW storage systems (Ahmed et al., 2016) considering that it needs to be available for use when requested. Fuentes et al. (2018) highlighted the importance of having relevant information about DHW withdrawal profiles and to consider them as a foundation on designing novel control strategies for DHW storage tanks. In addition to having representative load profiles, it is also important to model the thermal behavior of DHW tanks, so that it is possible to simulate these new control strategies, especially on different temporal resolutions. Currently there is an abundance in created DHW models, e.g. (Paull et al., 2010; Baeten et al., 2016; Jack et al., 2018) with different levels of detail (e.g. is stratification taken into account) (Jack et al., 2018). Additionally, there are also lot of different control strategies already created for utilizing DHW in DSM programs (Atikol, 2013; Jack et al., 2018; Paull et al., 2010). As there starts be more interest towards smart houses and full building control on smaller time resolutions, it is important to develop DHW models that can be integrated with other building-level models to ensure the potential for developing full building level control schemes and enable better participation to DSM programs.

1.2 Aims and objectives

The aim of the work is to create a DHW model with electric heating to be integrated into an existing Markov-Chain smart house model[1] (Louis et al., 2016) and study the impacts of various DSM programs to the electricity load profile of the heating of DHW tank. Therefore, the interoperability of the created DHW and existing smart house models need to be kept in mind during the creation of the model. Also, load shifting and other DSM tools are tested

---

[1] (`https://github.com/jeanlouisnico/SBuM`)

Inputs

Outputs
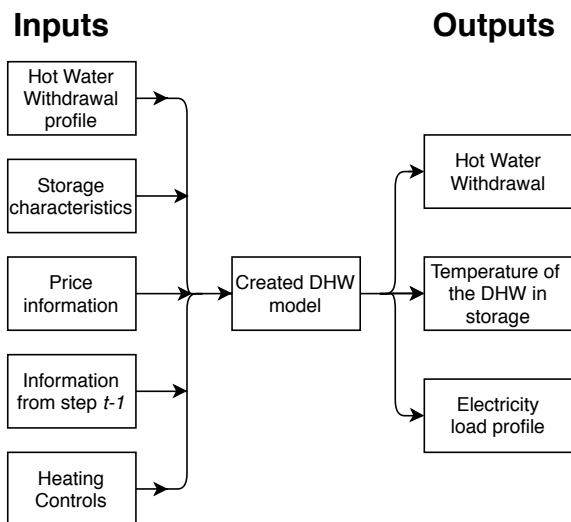
Figure 1. Basic description of the DHW model

Figure 2. Mean hourly draw-off profiles of the created 3 different 60 min load profiles (Default, 4 Categories option from Jordan and Vajen (2001) and load profile created from standard EN 12831-3).

to investigate the flexibility of electrically heated hot water storage tank. Finally, the results from the smart house model (Louis et al., 2016) and the created DHW model are investigated and compared on their load profiles as well as the potential to integrate the models.

2 Methodology

The created DHW model includes three separate parts: draw-off profile of hot water from the storage tank, temperature of hot water in the tank and heating schedules of the DHW. These parts allow modeling the behavior of the DHW tank to determine the electricity load profile and apply the control methods for various DSM programs. The model is created to be integrated in an existing smart house model (Louis et al., 2016) to study the electricity load management and potential DSM methods on building-level. Moreover it should provide more information about the building's thermal behavior and append the model's thermal part (Pulkkinen et al., 2019), while eventually enabling a more thorough control of electric heating in the building. There is also an on-going work to develop the smart house model to operate on a temporal resolution of 30, 15, 1 minute and even in 10 s timesteps. Therefore, it is important to follow this development to ensure the interoperability of these models.

The basic principle of the DHW model is presented in Figure 1. Basically, the model includes information on hot water demand, storage characteristics, price information and heating controls as separate inputs, while containing information about the previous simulation step (DHW temperature, heating load, etc.) for the use of the model. This provides the wanted outputs on the hot water withdrawal, temperature of the hot water and electricity load profile.
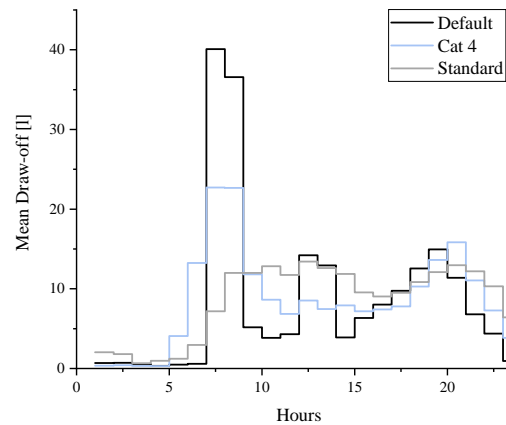
2.1 Hot Water Demand Profile

The hot water demand profiles for the model were created with using *DHWcalc* program, which creates realistic DHW load profiles using statistical methods on various temporal resolutions (Jordan and Vajen, 2001). Therefore, it was selected as a suitable tool for testing the created DHW model and its capability in running simulation on the created 60, 30 and 15 min timesteps. To ensure the capabilities of the DHW model, 3 different load profiles were created for each timestep, all for the length of one year. The first profile is the default profile created with default values from the program, while the second profile is created by inserting the hourly test profile for single-family houses from standard EN 12831-3 (2017) as step-function to the program. The third profile is created with the values from the 4 different categories option from the software, presenting default distribution per IEA-Task 26. The average daily DHW draw-off is 200 litres in each scenario and the mean hourly draw-offs for the created 60 min profiles are presented on Figure 2. The created profiles seem to have different characteristics as the Default load profile varies more on hourly basis and creates higher peaks than the other 2. Conversely, the load profile by the values from the standard has more balanced consumption during the day and has lower peaks. Additionally, the average 30 and 15 min draw-off profiles are presented in Figure 3, showing their different characteristics, and importance in simulating different time resolutions. In all cases, there were high morning peaks and 2 lower demand peaks, one around midday and the second in the evening.

2.2 Hot Water Tank

The model for DHW tank is created based on standards EN 15316-5 (2017) and EN 12831-3 (2017), and it is used to calculate the temperature of the hot water in the tank, stand-by heat losses and energy demand by utilizing energy balance method. The simplified calculation procedure with single volume and constant temperature
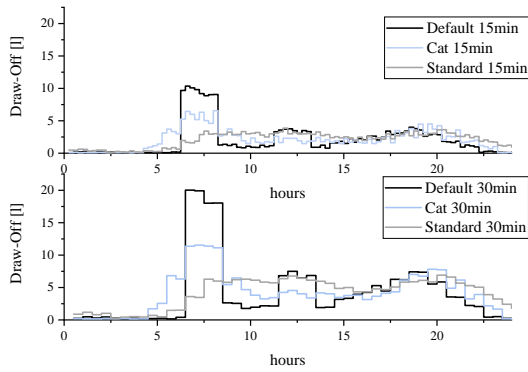
Figure 3. Mean draw-off profiles of the created 3 different 30 and 15 min load profiles

Figure 4. The calculation procedure for the created DHW model

throughout the tank was selected as the tank is assumed to be constantly mixed having therefore a uniform temperature profile throughout the tank.

The average temperature of the water T_{DHW} in the DHW tank is calculated with Equation 1:

$$T_{DHW} = T_{DHW,0} + \frac{P_{DHW} \times t_{ci} - Q_{DHW,out} - Q_{sto,ls}}{\rho_w \times C_{p;w} \times V_{sto;tot}} \quad (1)$$

where $T_{DHW,0}$ is the water temperature from last simulation step [°C], P_{DHW} is the heating power [W], t_{ci} is the calculation interval [h], $Q_{DHW,out}$ is the energy amount withdrawn from the tank [Wh], $Q_{sto,ls}$ is the stand-by heat loss from the tank [Wh], ρ_w is the density of water [$\frac{kg}{l}$], $C_{p;w}$ is the specific heat capacity of water [$\frac{Wh}{kg \times C}$] and $V_{sto,tot}$ is the total volume of the DHW storage tank [l] (EN 15316-5, 2017). Heating power for the DHW tank is determined according to the DHW heating part of the model, while energy withdrawn $Q_{DHW,out}$ from the DHW tank is calculated with Equation 2 and the stand-by energy loss $Q_{sto,ls}$ is calculated with Equation 3:

$$Q_{DHW,out} = V_{d,t} \times C_{p;w} \times \rho_w \times (T_{DHW,ref} - T_{in}) \times t_{ci} \quad (2)$$

$$Q_{sto,ls} = f_{bac,acc} \times f_{dis,ls} \times H_{sto,ls} \times (T_{DHW} - T_{amb}) \times t_{ci} \quad (3)$$

where $V_{d,t}$ is the volume of the draw-off at time t [l], $T_{DHW,ref}$ is the reference hot water temperature needed to create correct tab water temperature [°C] T_{in} is the temperature of the inlet water to the tank [°C], $f_{bac,acc}$ is the weighting factor for control and size of the tank [-], $f_{dis,ls}$ is weighting factor to correct the additional heat losses due to thermal bridges from the connecting pipes [-] and T_{amb} is the ambient temperature [°C] (modified from EN 15316-5 (2017) and EN 12831-3 (2017)).

2.3 Heating Demand

The basic principle for calculating the heating demand of the DHW tank and the resulting hot water temperature is presented in Figure 4. The model starts with calculating the temperature of the hot water in the tank if no heating

would be used. That information, among with the other inputs, is then delivered to the control part of the model to determine the heating profile of the tank. After that, the temperature of the hot water in the tank is calculated according to the heating power assigned to the tank from the previous step.

2.4 Heating Control

The heating of the DHW tank is determined in a separate part of the model according to the rules of the selected heating scenario. There are currently 4 different heating scenarios available on the model, which each has their own control mechanism and rules. The available scenarios are *Constant Temperature Set-point*, *On-Off Control*, *Time-of-Use Control* and *Linear Optimization*, of which the control rules of the first three are presented in Figure 5. All the scenarios take into account upper and lower temperature limits (T_{min} and T_{max}) as well as maximum heating power (P_{max}) as constrains in determining the heating profile. The lower temperature limit T_{min} is considered to be the minimum healthy temperature of water to prevent the growth of *Legionella* bacteria. All these variables can be determined by the user allowing the possibility to study the impact of them to the heating profile.

2.4.1 Constant Temperature Set-point

The first studied heating scenario is *Constant Temperature Set-point*, which control options aim in keeping a steady temperature inside the DHW tank. The rules for the controller are explained in Figure 5a) where the temperature set-point is called T_{min} and the maximum heating capacity P_{max} is considered as the second constrain.

2.4.2 On-Off Controller

The second heating scenario is *On-Off Controller*, which represents a control relay with 2 temperature set-points. The control rules for this scenario are presented in Figure 5b) and it operates only with full heating power until the upper temperature set-point T_{max} is reached, after which the heating only turns on after the hot water temperature drops under the lower temperature set-point T_{min}. In case the temperature set-points are reached in the on-going simulation time-frame, the power needed to meet them is
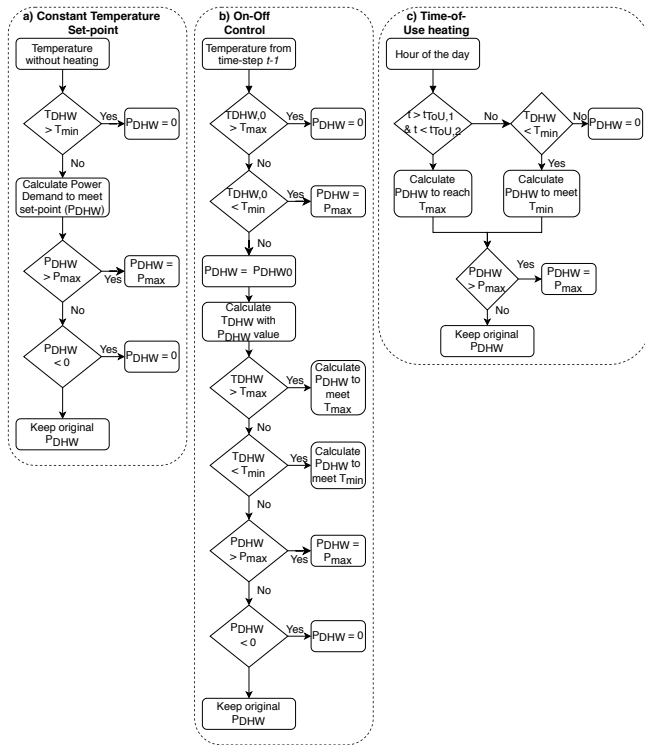
Figure 5. Presentation of the used rules in Constant (a), On-Off (b) and ToU (c) scenarios

calculated by solving P_{DHW} from equation 1 when T_{DHW} is $T_{DHW,min}$ or $T_{DHW,max}$ depending on the heating period, and the resulting power P_{DHW} is considered as the heating power for the time-step. This is a simplified representation of the real world On-Off controller as here the temperature set-points are always reached in the simulation time-steps (e.g. full hour) whereas in real-world the temperature could have already dropped under or increased over the temperature set-points when the measurement for the time-step is conducted. Therefore, this type of On-Off controller would be required to have an embedded smartness in it that would make the system meet the temperature set-point exactly at the simulation time period.

2.4.3 Time-of-Use Heating

The third heating scenario is *Time-of-Use Heating*, which participates in DSM by optimizing heating with a ToU price signal and using water as an energy storage medium. Generally, ToU represents a scenario where electricity price is divided to high and low price periods according to the times of high and low electricity demand periods, guiding electricity users to consume electricity during the lower price times. This is also the used rule in the heating scenario presented in Figure 5c). The main idea is that hot water is heated to and kept on the upper temperature set-point on the time of the lower electricity price, and not heated during the higher price period unless the hot water temperature drops under the lower temperature set-point, after which it is only heated up to keep the hot water temperature at the lower temperature set-point.

2.4.4 Linear Optimization

The fourth studied heating scenario is *Linear Optimization* of the electrically heated hot water tank according to hourly RTP signal and heat demand. This is another type of DSM action where the electricity load profile is varied according to the real time situation of the electricity network while maintaining an acceptable hot water temperature in the storage tank to ensure comfort and health of the user. The optimization goal is to minimize the electricity cost (Equation 4) with the constrains from Equation 5.

$$min \sum Cost_{1-n} = E_{DHW,1-n} \times Price_{1-n} \quad (4)$$

$$Constrains: T_{min} \leq T_{DHW} \leq T_{max}; 0 \leq P_{DHW} \leq P_{max} \quad (5)$$

where $Cost_{1-n}$ is the cost from electricity usage from timesteps *1-n* [€], $E_{DHW,1-n}$ is the energy used in timesteps *1-n* [MWh] and $Price_{1-n}$ is the real time price of electricity during the same timesteps [€/MWh]. The T_{DHW} is calculated with Equation 1, to which also forecasted hot water demand is added. The optimization period *n* can be determined by the user, but values of over 4 timesteps start to become slow computing-wise. For instance, linear optimization simulation with optimization period of 4 timesteps takes 495s for hourly values for a year (8760 simulation steps), whereas optimization periods of 3 and 5 timesteps take 377s and 663s respectively for the same input data. This would still allow optimizing the system on real-time with the optimization period of 4 or 5 timesteps, but the difficulty in applying it to practice comes from the uncertainty in determining the draw off profile from the DHW tank as estimating the energy draw off correctly is vital to the correct optimization.

2.5 Inputs

The inputs values for testing the DHW behavior and DSM potential are defined on Table 1 underneath from the Decree 1010/2018 [2], and the international standards EN 15316-5 (2017). Some input values vary depending on the heating scenario, while others stay the same during all the simulations.

The minimum requirement for the DHW in EN 15316-5 (2017) is considered as 55 °C, which is also the lower temperature setpoint used in the simulation, except for the *Constant Temperature Set-point* scenario where a temperature set-point of 60 °C was used to balance the otherwise occurring lower average temperature. The hourly Elspot price for Finnish transmission area in 2016 (Nord Pool, 2018) was selected as the price signal and the forecasted Hot Water Demand was created based on the test profile for single-family houses from EN 12831-3 (2017). The stand-by heat losses were calculated according to EN 15316-5 (2017), EN 60379 (2004) and EN 50440 (2015) by assuming a vertically adjusted and electrically heated DHW tank.

Table 1. Input values for the simulation (EN 15316-5, Decree 1010/2018)

Variable	Value
Hot Water Tank Volume	300 l
Max Power	3 kW
Inhabitants	4
Average DHW demand	50 $\frac{l}{inh}$
Inlet Water Temperature	10 °C
Ambient air temperature	16 °C
Default T_{DHW} and T_{in} difference	45 °C
T_{min} for Constant Temp Set	60 °C
T_{min} for On-Off, ToU and Optimization	55 °C
T_{max} for On-Off	65 °C
T_{max} for ToU and Optimization	85 °C
ToU heating hours	22-6
LP optimization period (timesteps ahead)	4

3 Results

The results firstly investigate the suitability of integrating the created model to the existing Markov-Chain smart house model by presenting the capability of operating in similar time resolutions and providing useful and robust results. Secondly, the behaviour of the DHW heating controls are examined in the 4 hourly heating scenarios. After that, the flexibility of DHW heating is investigated and the effectiveness of various control and DSM mechanisms are presented. The final part shows the combined effect of the two models in order to present the importance in integrating the smart house and DHW models.

3.1 Robustness of the results

The robustness of the results is looked into by investigating similar thermal behaviors of the DHW tank and by comparing annual results to each other when using *Constant Temperature Set-point* scenario. The annual results from the simulations are available at Table 2.

The results from Table 2 indicate that there are very little differences on annual level on the results with DHW

Table 2. Results from the Constant temperature heating scenario

	Total Heating [MWh]	Average Temperature [°C]	Total Costs [€]
Default 60min	7.19	59.74	251.44
Default 30min	7.20	59.75	251.28
Default 15min	7.19	59.73	251.25
Cat 60min	7.20	59.78	245.42
Cat 30min	7.19	59.66	244.98
Cat 15min	7.18	59.60	244.84
Standard 60min	7.22	60.00	245.08
Standard 30min	7.21	60.00	244.78
Standard 15min	7.21	59.97	244.48

60, 30 and 15 min withdrawal profiles. As the *DHWcalc* uses statistical methods in withdrawal profile creation, it is understandable that there are small differences on the results, but this could be considered as noise in the simulation as well.

3.2 Behaviour of DHW heating controls

This sections investigates the behaviour of the created DHW heating controls by looking into the DHW heating power frequency and cumulative frequency distributions on hourly heating scenarios. These results are presented in Figure 6 showing firstly the similarities between *Constant Temperature Set-point* and *Linear Optimization* scenarios, which rarely use maximum heating power and have more lower power consumption hours. This resemblance comes from the short optimization time in *Linear optimization* scenario, which makes both of them aim for lower water temperature in the DHW tank. Secondly, Figure 6 shows similarities of *On-Off* and *ToU* scenarios, which are more prone to have more maximum and minimum power periods and less intermediate power consumption. This relates to their control rules, as they both utilize maximum power until they reach the upper temperature limit, after which they do not use heating until they have reached the lower temperature set-point. The difference between these 2 control strategies lie in the maximum heating time-periods. *On-Off* scenario starts the maximum heating immediately once it has reached the lower temperature limit while *ToU* control keeps the DHW tank at the lower temperature point until the start of the cheaper night-time heating period. This is also the reason of their variations in 0-0.5 kW and 0.5-1.0 kW power consumption.

3.3 Flexibility of DHW

This section investigates the flexibility of DHW and the viability of few control mechanisms to provide DSM. First, the reference values for the flexibility are visible on Ta-
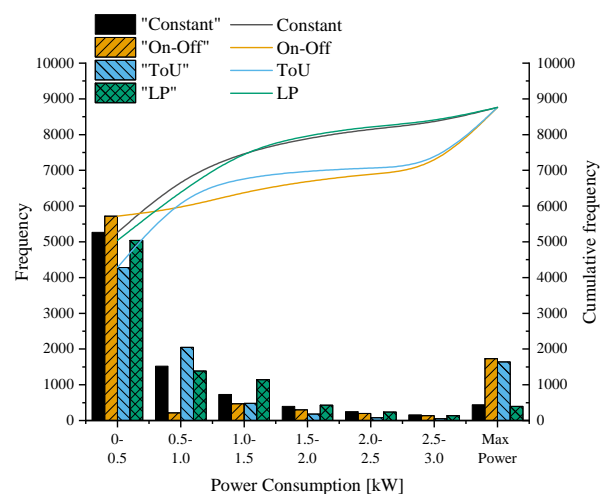
Figure 6. Frequency (bar) and cumulative frequency distributions (line) of the DHW heating scenarios in hourly time-scale

Table 3. Annual results from On-Off, Time of Use and Linear optimization scenarios

	On-Off Control			Time of Use			Linear Optimization		
	Total Heating [MWh]	Average Temperature [°C]	Total Costs [€]	Total Heating [MWh]	Average Temperature [°C]	Total Costs [€]	Total Heating [MWh]	Average Temperature [°C]	Total Costs [€]
Default 60min	7.27	60.70	255.22	7.88	68.44	222.89	6.87	55.52	230.55
Default 30min	7.23	60.22	253.84	7.81	67.53	219.60	6.83	55.07	234.04
Default 15min	7.22	60.03	254.29	7.80	67.50	219.26	6.82	54.97	236.58
Cat 60min	7.27	60.62	249.50	7.96	69.53	219.70	6.87	55.53	224.47
Cat 30min	7.23	60.11	248.58	7.88	68.44	216.28	6.83	55.01	227.77
Cat 15min	7.21	59.89	248.80	7.87	68.35	215.78	6.82	54.93	230.37
Standard 60min	7.28	60.87	248.44	7.94	69.23	218.87	6.88	55.74	223.93
Standard 30min	7.25	60.41	246.88	7.90	68.73	216.80	6.84	55.17	226.84
Standard 15min	7.23	60.21	245.85	7.90	68.72	216.42	6.83	55.08	229.25
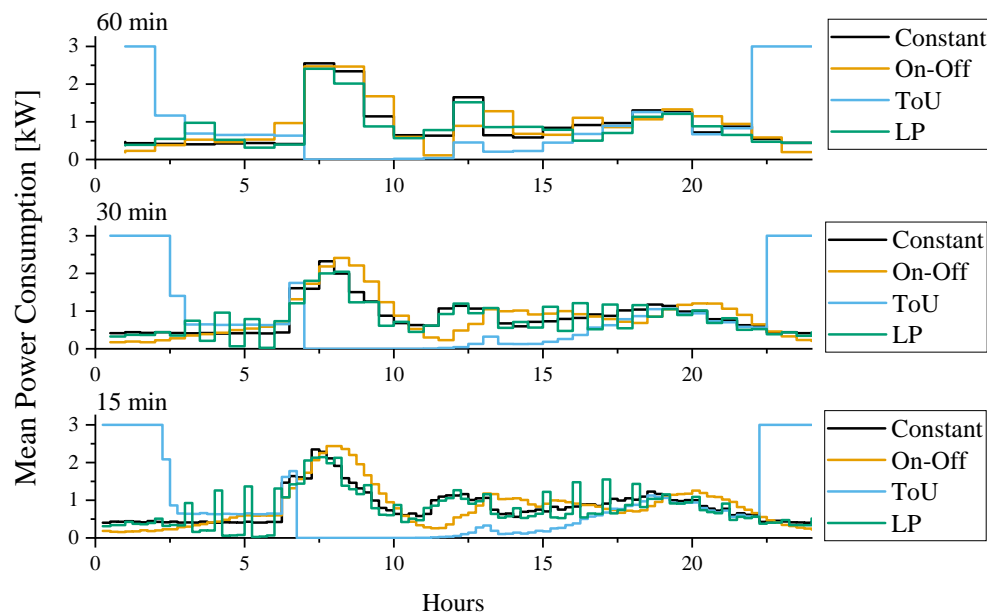
Figure 7. Comparison of the electricity load profiles of different heating scenarios with 60, 30 and 15 min time resolution.

ble 2. The annual values for the other heating scenarios are available on Table 3 and the average hourly power for the DHW tank in Figure 7. The results from Table 3 indicate that changes between the time resolutions are generally small generating difference of less than 2% in heat demand or average hot water temperature, while the difference between cost can be slightly higher varying from -1.8 to 2.6%.

Comparing different scenarios to each other, reveals the flexibility of the DHW source and the potential to participate in DSM programs. On annual level, the cheapest option seems to be night time charging with *ToU* scenario, which is 13.5-16% cheaper than *On-Off Control* scenario and 2.1-7.3% cheaper than linear optimization scenario, while having 7.4-15.9% higher electricity consumption and 10.8-24.8% higher average hot water temperature. This reveals the potential of storing energy to

DHW tank during the night-time. The higher hot water temperature is also beneficial in terms of comfort as *ToU* scenario was the only one where the temperature of DHW tank was always over the lower temperature set-point, being able to provide enough hot water to use.

Looking into the average hourly heating loads from Figure 7 the night time charging of the DHW tank is clearly visible as it has higher electricity consumption during night hours and low consumption during the day. Similarly, the morning peak in the draw-off of hot water creates a heating peak in all scenarios, but *ToU*. Otherwise, *Constant temperature set-point* and *Linear optimization* scenarios had pretty similar load profiles, only difference being the occurring peaks with *Linear optimization* scenario in shorter time resolutions. These peaks will occur as the electricity price will remain the same for the whole hour, so the controller aims at increasing the temperature
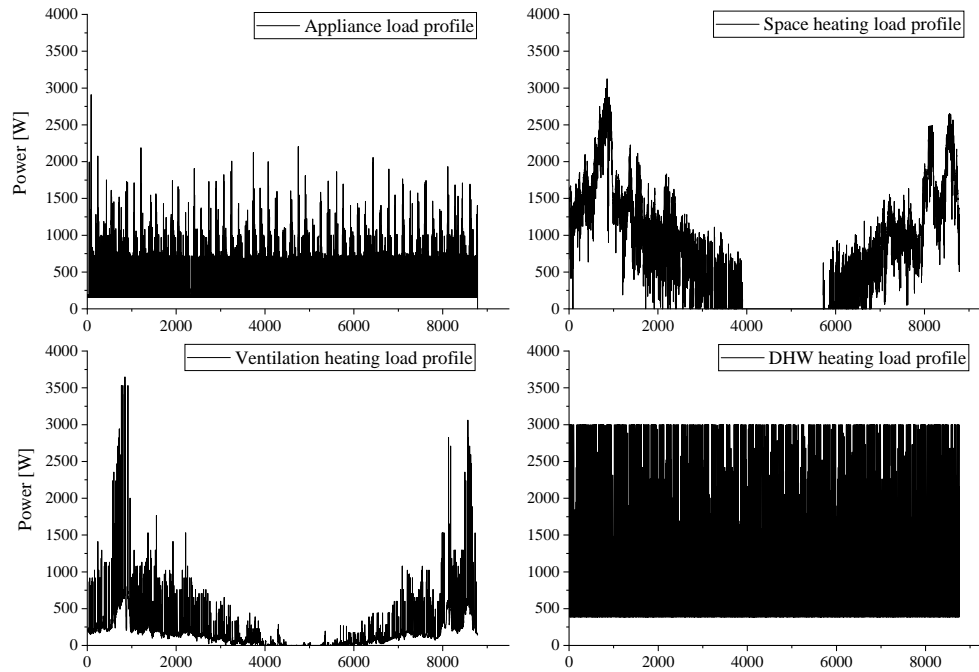
Figure 8. Comparison of various building loads

of the hot water on the earliest available time slot. *On-Off control* then had a delay on the heating compared to *Constant* and *Linear optimization* scenarios, as with *On-Off controller* the temperature of the water needs to drop under the lower temperature set-point before it applying the maximum heating power.

3.4 Load profiles

This part investigates the importance of integrating DHW model into the existing smart house model (Louis et al., 2016), with which it is possible to simulate appliance's load profiles, thermal behavior of the building and electric space and ventilation heaters. An example simulation was conducted with the smart house model and the achieved load profiles were compared to the load profile of the DHW heater (hourly constant temperature scenario) in Figure 8 and annual values in Table 4. The simulated building was a detached house, with building characteristics similar to 2018 building type from (Pulkkinen et al., 2019) with randomly generated appliances.

Results from Table 4 show that the DHW heating has the highest electricity consumption in the example building, and from Figure 8 it becomes evident that DHW has

highly varying, but seasonally rather constant load profile, similarly to appliances. Conversely, space and ventilation heating loads tend to require a lot of power during winter, but have low power demand in summer. This means that all these loads have their different characteristics and load profiles which while accumulated would have multiple impacting sources to the building's energy management.

4 Conclusions

The aim of this work was to create a functional, fully mixed DHW tank model to be integrated in an existing Markov-Chain smart house model and to investigate the flexibility of DHW load on 60, 30 and 15 min time resolutions and with 4 different heating controls. The behavior of the model was tested with using DHW withdrawal profiles created with *DHWcalc* program. The results indicate that the model was able to behave robustly with a *Constant Temperature Set-point* scenario in all tested time resolutions, showing little difference between their simulation results.

Comparing the results of different heating control scenarios from Table 3 the flexibility of the electricity consumption with DHW tank is visible in the *ToU* scenario results as it has the lowest costs and highest total heating energy consumption of the studied scenarios. Therefore, with *ToU* scenario it is possible to use DHW tank as an energy storage and charge it during the lower night time electricity prices. Yet, this behavior is contradictory as it reduces the efficiency of the system from electricity consumption point-of-view. The low performance from the *Linear optimization* scenario compared to *ToU* was likely related to the short optimization time and mis-

Table 4. Annual results from the example simulation

	Annual Energy Consumption [MWh]
Appliances	2.91
Electric space heating	7.08
Electric ventilation heating	1.50
Electric DHW heating	7.19

matches between the forecasted and actual energy withdrawals. These did not allow the optimized heater to utilize the thermal inertia of the DHW tank and lower electricity prices during the night as it did not expect a withdrawal from the tank early enough. Similarly, the difficulty in knowing the correct withdrawal of energy reduced the accuracy of the optimization. The final result showed the importance in integrating the DHW model with smart house model for allowing testing of new control mechanisms and different aggregated load profiles. Furthermore, in an integrated model all the separate electricity loads can be controlled together.

The future direction of the work is to finalize the integration of the smart house and DHW models, and start developing combined control mechanisms for them. Also the current DHW model should be developed to operate in lower time resolutions to allow also short-term power network testing, as well as expanded to include temperature stratification for increased accuracy and different DHW tank options. Consideration of improving the optimization with aggregated load profiles should be investigated as well. Also renewable energy generation and control mechanisms related to it should be added to the model.

Acknowledgement

References

K. Ahmed, P. Pylsy, and J. Kurnitski. Hourly consumption profiles of domestic hot water for different occupant groups in dwellings. *Solar Energy*, 137:516–530, 2016. doi:10.1016/J.SOLENER.2016.08.033.

U. Atikol. A simple peak shifting DSM (demand-side management) strategy for residential water heaters. *Energy*, 62:435–440, 2013. doi:10.1016/j.energy.2013.09.052.

B. Baeten, T. Confrey, S. Pecceu, F. Rogiers, and L. Helsen. A validated model for mixing and buoyancy in stratified hot water storage tanks for use in building energy simulations. *Applied Energy*, 172:217–229, 2016. doi:10.1016/j.apenergy.2016.03.118.

EN 12831-3. Energy performance of buildings. Method for calculation of the design heat load. Part 3: Domestic hot water systems heat load and characterisation of needs, Module M8-2, M8-3. European Standard, 2017.

EN 15316-5. Energy performance of buildings. Method for calculation of system energy requirements and system efficiencies. Part 5: Space heating and DHW storage systems (not cooling), Module M3-7, M8-7. European Standard, 2017.

EN 50440. Efficiency of domestic electrical storage water heaters and testing methods. European Standard, 2015.

EN 60379. Methods for measuring the performance of electric storage water-heaters for household purposes. Standard, 2004.

E. Fuentes, L. Arce, and J. Salom. A review of domestic hot water consumption profiles for application in systems and buildings energy performance analysis. *Renewable and Sustainable Energy Reviews*, 81:1530–1547, 2018. doi:10.1016/j.rser.2017.05.229.

M. W. Jack, K. Suomalainen, J. J. W. Dew, and D. Eyers. A minimal simulation of the electricity demand of a domestic hot water cylinder for smart control. *Applied Energy*, 211:104–112, 2018. doi:10.1016/j.apenergy.2017.11.044.

S. Ø. Jensen, A. Marszal-Pomianowska, R. Lollini, W. Pasut, A. Knotzer, P. Engelmann, A. Stafford, and G. Reynders. IEA EBC Annex 67 Energy Flexible Buildings. *Energy and Buildings*, 155:25–34, 2017. doi:10.1016/J.ENBUILD.2017.08.044.

U. Jordan and K. Vajen. Influence of the DHW load profile on the fractional energy savings: A case study of solar combisystem with TRNSYS simulations. *Solar Energy*, 69:197–208, 2001. doi:10.1016/S0038-092X(00)00154-7.

J-N. Louis, A. Caló, K. Leiviskä, and E. Pongrácz. Modelling home electricity management for sustainability: The impact of response levels, technological deployment & occupancy. *Energy and Buildings*, 119:218–232, 2016. doi:10.1016/j.enbuild.2016.03.012.

N. Lu. An Evaluation of the HVAC Load Potential for Providing Load Balancing Service. *IEEE Transactions on Smart Grid*, 3(3):1263–1270, 2012. doi:10.1109/TSG.2012.2183649.

Nord Pool. Historical Market Data, 2018. URL https://www.nordpoolgroup.com/historical-market-data/. [Accessed 07.07.2020].

L. Paull, H. Li, and L. Chang. A novel domestic electric water heater model for a multi-objective demand side management program. *Electric Power Systems Research*, 80(12):1446–1451, 2010. doi:10.1016/j.epsr.2010.06.013.

J. Pulkkinen, J-N. Louis, and E. Pongrácz. Utilizing flexibility of electric heating in demand side management programs in Finland in 2050. In *Proceedings of 11th International Conference on Applied Energy, Part 3, Sweden, 2019 (ICAE2019)*, volume 4, 2019.

J. Salpakari and P. Lund. Optimal and rule-based control strategies for energy flexibility in buildings with PV. *Applied Energy*, 161:425–436, 2016. doi:10.1016/j.apenergy.2015.10.036.

US Department of Energy. Benefits of Demand Response in Electricity Markets and Recommendations for Achieving Them, A report to the United States Congress Pursuant To Section 1252 of Energy Policy Act of 2005. Technical report, 2006. URL https://www.energy.gov/oe/downloads/benefits-demand-response-electricity-markets-and-recommendations-achieving-them-report. [Accessed 07.07.2020].

Predicting Thermal Losses in a Low-Energy Building

Jakob Pfaff[1]   Elisabet Syverud[2]   Dag A. H. Samuelsen[3]

Department of Science and Industry Systems, University of South-Eastern Norway, Kongsberg, Norway
1) pfaff901@gmail.com 2) Elisabet.Syverud@usn.no 3) Dag.Samuelsen@usn.no

Abstract

Energy management of small-scale renewable energy systems (microgrid) requires control of the energy consumption. Commercial buildings are sized using energy consumption criteria from the building standards. The energy consumption of a building depends on a number of dynamic factors, including thermal loss, building climate control, and building utilization. This paper presents a simple model for predicting the thermal loss of a building. The basic simulation module is a single room model where the outer walls and windows are exposed to the ambient conditions without sun influx. We model the thermal loss based on building design data and validate the model using operational data for an actual low-energy building. The model prediction accuracy is within +/- 1°C for up to 7 days when predicting thermal losses in the building construction.

Keywords: energy management, powerhouse, thermal loss prediction, Matlab

1   Introduction

A powerhouse is a building that produces more electric power than it consumes during the building lifecycle, including construction and decommissioning/disposal. Sustainability conscious real estate asset owners are interested in the concept and Norway has seen a rise in the construction of powerhouses in the past 4-5 years.

Powerhouses have large solar PV systems installed at the roof. At its maximum, the solar PV systems generate significantly more energy than consumed by the building at any time. In addition, the maximum power production occurs in the summer season when the heating demand is at its yearly minimum. Heating is the main energy consumer of commercial buildings. Thermal losses drive the heating demand. In our research, we explore models for predicting thermal losses in a low-energy building. The model will be used as part of a digital twin of the building's energy system. Our overall goal is to develop practices for energy management of powerhouses.

The digital twin is a dynamic virtual model of a system that incorporates business, contextual, and sensor data from physical systems into the virtual system model (Madni et al., 2019). A true digital twin is different from other models because it includes the specific instance(s) that reflects the characteristics of the physical twin, in real time. Madni et al. (2019) sees the digital twin as promising technology, particularly in the building and real estate industries, because of real time access to system data that are essential for energy management (Madni et al., 2019).

Research papers suggest different approaches for assessing thermal performance of buildings. (Kildsgaard et al., 2013; Makaka, 2015; Rohdin et al., 2014) are relevant examples for low-energy buildings and passive houses. Both physical models and machine learning models are available in literature. Mendes et al. (2001) give a first principle (physical model) of the thermal loss. In the digital twin, we need a model that integrates the virtual and physical world. We need to base our model on simple approaches because calculation speed will be an issue due to the large amount of data in the system and the need for real time monitoring and prediction.

Lie (2019) defines the hybrid model as a combination model. A hybrid, data driven model improves the fit between an imperfect physics-based model and (inherently limited) experimental data. The hybrid model is particularly interesting for digital twin applications. To ensure fast response of the digital twin, it is essential that the model is as simple as possible.

This paper outlines a simple, first principle based model for calculating the thermal losses of a low-energy building. The model predicts the energy losses from the building when exposed to varying ambient conditions. We base our model on the design assumptions of the building and we validate the model using operational data from the building. We ask in our research how well a simple energy balance model can predict the thermal losses of a real building if we base the model on the design parameters of the building.

2   Vestsiden Middle School Case

We use Vestsiden middle school in Kongsberg as our case. Kongsberg Kommunale Eiendom (KKE) completed the building construction in August 2019. The school is a low-energy house. Power and heat is supplied through geothermal heat and solar power systems. The national electric grid connects to the building. Solar power is produced by 1054 solar PV panels installed at the roof. Excess energy is stored in a battery and consumed at night. KKE plans to install an electrolyzer and hydrogen storage system for seasonal storage. A fuel cell will generate electric power from the hydrogen during winter. There are currently no guidelines for dimensioning and operation of a hydrogen loop in a hybrid renewable energy system for buildings (Bredesen, 2019).

2.1 Construction Standards

The design of commercial buildings in Norway comply with NS-EN ISO 52000 (energy performance standard) and NS3701 (passive house standard). The standards allow for assessment of the power demand of a building. The designers use the standards to assess if the building design complies with the regulations. The standards provide a specification basis for the products and construction elements.

Special purpose simulation programs assist the building designers in the calculation of energy consumption of the building at various ambient conditions throughout a typical year (SIMIEN, 2020).

The construction company adjusts the heat loops and ventilation flow rate so that the as-built building complies with the design intent. The building constructor sets the water valves and ventilation dampers during commissioning.

2.2 Real Time Measurements

The Supervisory Control and Data Acquisition System (SCADA) monitors and controls the building in multiple zones. More than one thousand instruments are installed in the building. The data sample rate depends on the PLC, and is typically 1 minute. At a minimum, the controller measures air temperature, carbon dioxide level, and ventilation flow rate in each zone. The SCADA system sends the data to a database located at USN Campus Kongsberg.

We measured the actual air temperature in the room and the ambient temperature during the months of March and April 2020. The four weeks from 18th of March to 16th of April 2020 are particularly interesting. Due to the 2020 pandemic, the school was in lockdown and there were no people in the building during this time.

3   Design Basis and Assumptions

To predict the thermal losses and verify with actual room temperature measurements, we need to understand the energy balance of the building. The outside air temperature and the radiation from the sun affect the building, Figure 1. The basic simulation module is a single room model where the outer walls and windows are exposed to the ambient conditions.

We do not consider the effect of solar energy influx in this work. The room used for model verification is facing North-West with the horizon at 315 degrees.

Multiple effects influence the energy balance at a given time. Figure 2 shows a sketch of the contributors to the energy balance of a room. We combine the most relevant thermodynamic processes in our model: a) thermal losses of walls, windows and roof, b) thermal mass of furniture, and c) heat flow contributed through air ventilation, floor-heating, radiator heating, and people (users).

3.1 Walls, Windows, and Roof

The energy performance standard governs the building design. We use the same approach in our model and we use the overall heat transfer coefficient, U-factor, and heat storage capacities applied in the building design. The data are available from the building owner. With this approach, we do not need the building construction details like the composite wall design or insulation material. Using the design basis allows us a direct comparison with the design basis.
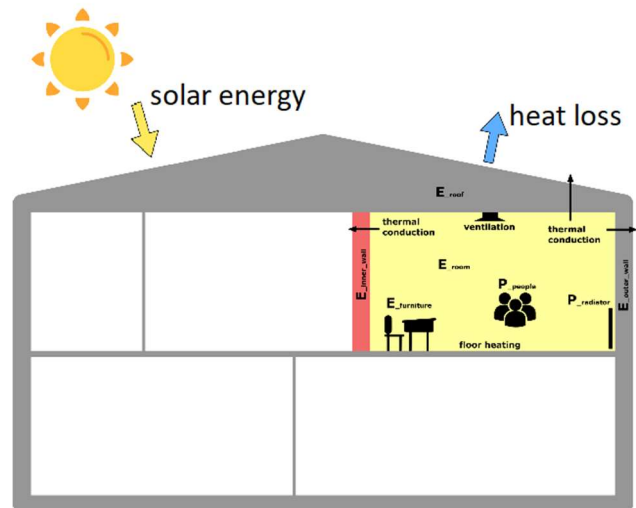
Figure 1. Thermal loss model of building
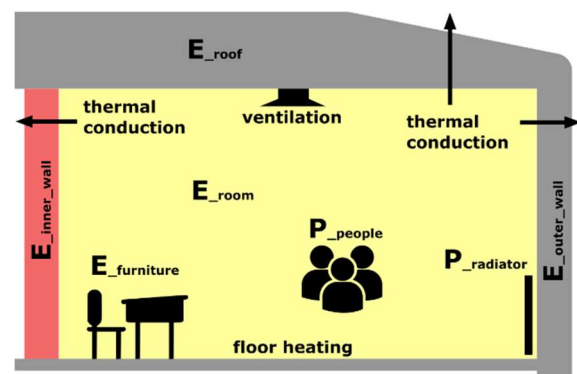
Figure 2. Energy sources and thermal losses in the room

The outer wall of the room is $22m^2$ of which 55% is window surface. The floor and roof surface is 63 m^2. We assume zero heat transfer across the inner walls of the building. There are no doors exposed to ambient air.

In our work, we use the following building design parameters:

Heat transfer coefficient (U-factor)

- Outer wall: 0.17 W/m^2K
- Roof: 0.13 W/m^2K
- Windows: 0.85 W/m^2K

Heat capacity (h):
- Outer wall: 12 Wh/m²K
- Roof: 10 Wh/m²K
- Window: 3 Wh/m²K

3.2 Furniture

Changes in the air temperature will change the furniture temperature. Depending on the furniture mass, this can cause a shift in the modeled air temperature compared with measurement.

For design basis, we assume 30kg of furniture per student (900 kg total for the room in this work). We assume a specific heat capacity of 1000J/kgK for the furniture.

3.3 Ventilation

The ventilation provides a flow of fresh air to the room. The airflow rate (mass based) at inlet to the room is the same as the airflow rate leaving the room (per mass). The control system measures the volumetric airflow rate. This is a rudimentary measurement based on the valve position. During normal use, the valve position is adjusted automatically based on the actual demand regulated by the temperature setting of the room and the air quality (local measurements of carbon dioxide).

The flow rate to the room is zero when the ventilation system is shut down and 1300 m³/hr at full flow. During normal operation, the control system schedule sets ventilation flow rate.

We assume air as an ideal gas with constant specific heat capacity (isochoric) of 717.3 J/kgK and a density of 1.2 kg/m³ for the air.

3.4 Floor-heating

The building floor-heating is installed in the basement, shared rooms, library and the main hall. Water circulates in piping embedded in the floor. A geothermal heat pump heats the water. Electric heaters allow for higher temperatures during the cold season. The water flow rate is per design and not adjusted after commissioning

There is no floor-heating in the room used in our model and the flow rate is therefore set to zero in this work.

3.5 Radiator

Most rooms have one radiator. Water circulates through the radiator and the heat is released through natural convection to the air. Each radiator can provide up to 1400W heat to the room. The actual heat release depends on the valve setting in the water loop. The water flow rate is not measured.

The energy balance calculated at the start condition of the simulation determines the heat input to the room. We use a constant radiator heat input of 290 W in this work, which is the initial state equilibrium of the room for the measurements used in here.

3.6 People

People using the room will release heat to the surroundings. The energy performance standard recommends a heat release of 12 W/m2 per person for design of school buildings. The effect of people is not included in this paper because we collected the building data for a time where there were no people in the building.

4 Thermal Model

We use MALAB Simulink, v.10.0 (R2019b) for the simulations. Our system is an open system with conservation of mass (air). From the first law of thermodynamics, assuming zero work, we find that the change in internal energy from the initial state 0 to the final state is equal to the change in heat (Q) in the system:

$$E_{air} - E_{air0} = \int Q \, dt \qquad (1)$$

where E_{air} is the energy of the air in the room and 0 indicates the initial state.

The net heat transfer to the room is the sum of heat flowing into or out of the system.

The air temperature at the initial state is calculated from the internal energy:

$$T_{room} = \frac{E_{air}}{cv_{air}*\rho_{air}*V_{room}} \qquad (2)$$

Where cv_{air} is the isochoric specific heat capacity of the air; ρ_{air} is the density of air, assumed constant; V_{room} is volume of the room), Figure 3.
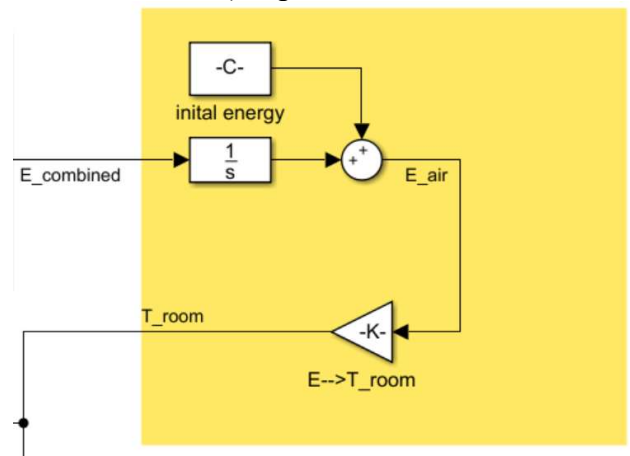
Figure 3. Room temperature calculation in MATLAB

4.1 Wall, Window and Roof

The wall is modelled by the thermal resistance, eq.3, Figure 4:

$$E_{wall} = \int Q_{wall} \, dt + E_{w0} \qquad (3)$$

We base the wall model on Fourier's law of heat conduction where the heat transfer through a material is proportional to the (negative) gradient in the temperature and to the area perpendicular to that gradient. The U-factor determines the heat transfer.

$$q_{conduction} = (T_{wall} - T_{outside}) * U_{wall} * A_{wall} \quad (4)$$

This heat flow does not directly interact with the room but reduces the temperature of the wall that again increases the heat flow through the convection model.

We include the heat (storage) capacity of the material in our model to allow for a delay in the development of the air temperature inside the room:

$$q_{wall} = (T_{room} - T_{wall}) * h_{wall} * A_{wall} \quad (5)$$

where , A_{wall} is the surface area of the outer wall.

This is as well the interface for this sub model to the central room model, as it transfers heat between the air in the room and the thermal mass of the wall.
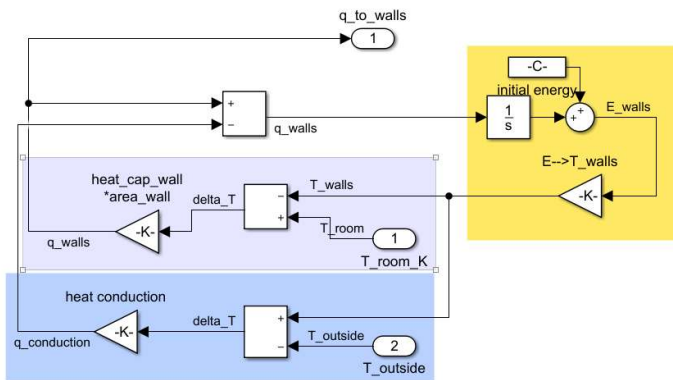
Figure 4. Wall model in MATLAB

The windows and roof models are analogous to the wall model with different properties as listed in Section 3.1.

We use room air temperature to connect the sub-models as seen in Figure 5. The air temperature calculation converges to steady state at each time step.

4.2 Furniture

The furniture model uses the thermal mass model since there is no sufficient data for a convection-based model and no heat conduction is taking place. See Figure 5.

4.3 Ventilation

The ventilation model uses volumetric flowrate $\dot{V}_{vent}$ and initial air temperature T_{vent} to determine the heat flow of the incoming air to the room:

$$q_{vent\_in} = T_{vent} * c_{air} * \rho_{air} * \dot{V}_{vent} \quad (6)$$

Figure 6 shows the model of the ventilation at the inlet to the room, the heating and the heat input from people.

Because air pressure inside the room is constant, the air mass flow from the room is the same as at the inlet. The heat flow of the outgoing air from the room becomes (air density assumed constant):

$$q_{vent\_out} = T_{room} * c_{air} * \rho_{air} * \dot{V}_{vent} \quad (7)$$

where the temperature of the outgoing air is the temperature of the bulk air temperature in the room. The model is shown in Figure 5 (purple box).
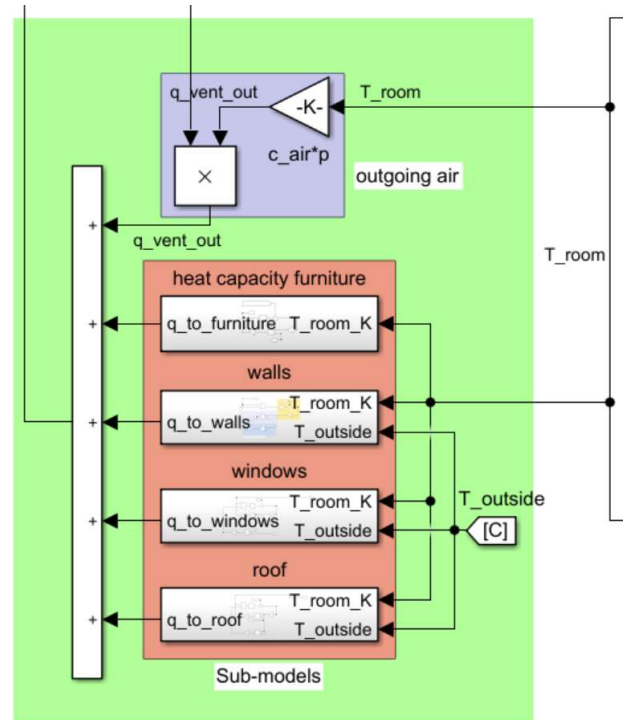
Figure 5. Thermal models connected by air temperature
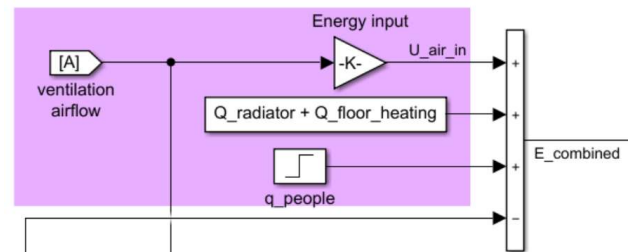
Figure 6. Ventilation flow model in MATLAB

4.4 Floor-heating

The floor-heating model is based on the difference in water temperature as measured at the inlet and outlet of each building zone. The water flow rate is constant.

$$q_{floor} = c_{water} * \rho_{water} * \dot{V}_{water} * (T_{in} - T_{out}) \quad (8)$$

Floor-heating is set to zero in this paper.

4.5 Radiator

We use a constant heat input of 290W from the radiator to the air; see Figure 6 and Section 3.5.

4.6 People

The heat released by the users is not included in this work; see Section 3.6.

5 Model Validation

We validate the model using actual data measurements from a room in the school building. We use two different data series:

1. Ventilation system fully closed (due to pandemic lockdown),
2. Ventilation system operated according to the standard building control scheme (with low flow during night).

The SCADA samples ambient temperature, room air temperature, and ventilation airflow rate at 1-minute intervals for both data series.

We use the measured ambient temperature and the ventilation airflow rate as input to the model and simulate the room air temperature. In the final digital twin, we will use a temperature forecast. The results presented in this paper are therefore assuming a perfect forecast.

Finally we compare the simulated air temperature to the measured air temperature, see Figure 7 for model.
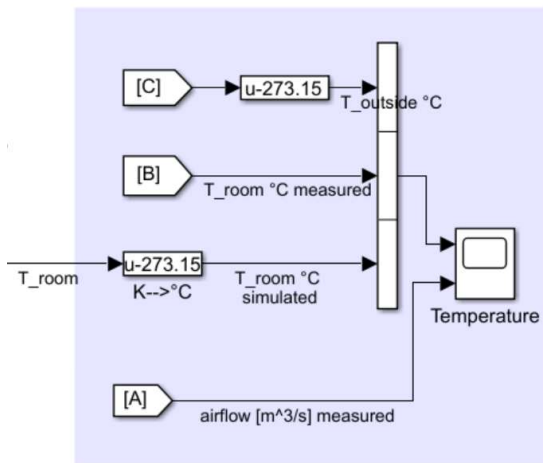
Figure 7. Room Temperature Validation

6 Results and Discussion

Figure 8 shows the simulated and measured room temperature for the case of the shutdown ventilation system. The ventilation airflow rate was zero during the time and the ambient temperature varies from -3°C to +14°C in diurnal cycles (14 days).

The thermal loss model predicts the response of room air temperature to the variation in ambient temperature. Figure 8 shows fair prediction accuracy with a maximum deviation of +/-1°C between the simulated and the measured room temperature. After 7 days, the simulation starts to deviate from the measured data and the prediction is less reliable.

The prediction accuracy remains high despite the coarse resolution in the ambient air temperature measurement used as input to the model. The ambient temperature instrumentation installed in the building is accurate. The reduced resolution and observed data discretization is due to data truncation in the

measurement system. We recommend improving this for the future.
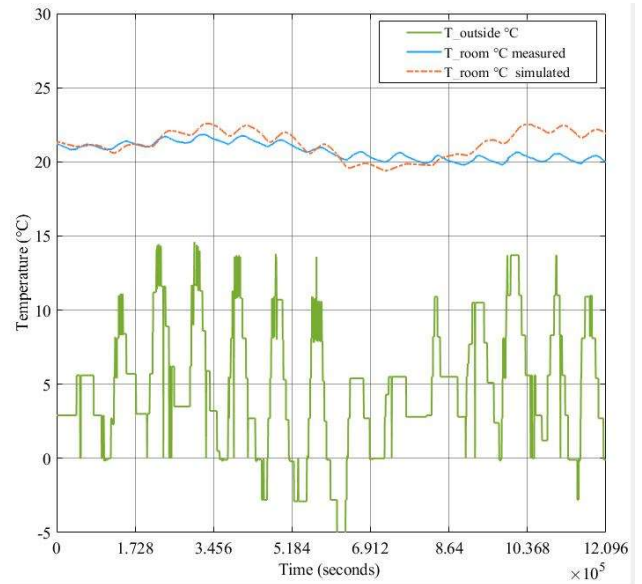
Figure 8. Room temp prediction with closed ventilation

Figure 9 shows the simulated and measured room temperature for the case with the ventilation system in normal operation. The ambient temperature varies from -3°C to +28°C in diurnal cycles (14 days). The ventilation airflow (Figure 10) is operating during daytime on the weekdays and fully closed during the weekend. The ventilation air temperature is constant for the full simulation.

The prediction accuracy of Figure 9 is reduced significantly compared to Figure 8. The model is not able to predict the behavior for more than 2 to 3 days. The simulation assumes a constant ventilation air temperature. In reality, the air temperature changes within the limits set in the building control system. This effect is not included in the simulation.
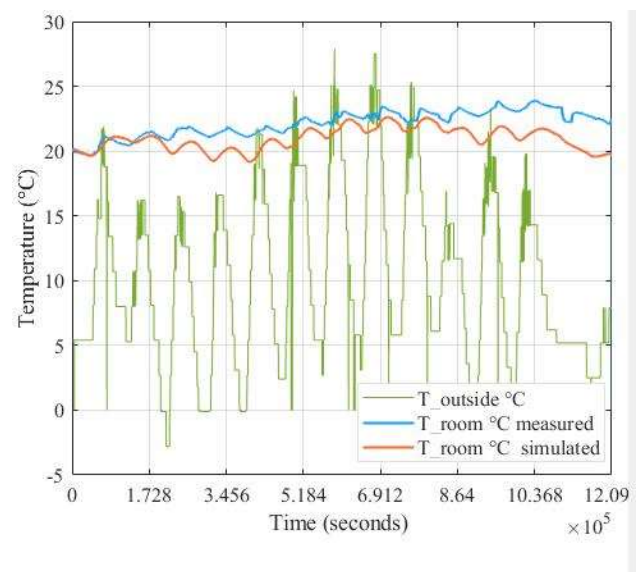
Figure 9. Room temp prediction, ventilation running in normal operation
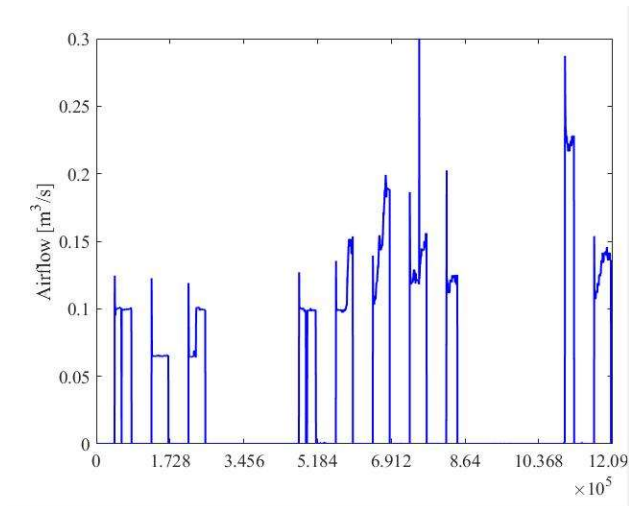
Figure 10. Ventilation air flow rate applicable to Fig 9

7 Conclusions and Further work

Low-energy buildings have significant heat loss to the surroundings. This causes diurnal temperature variation inside the building and drives the energy consumption. We search for a simple model to predict the thermal losses of a low-energy building. We will use the model to predict the energy consumption of the building for short time periods (up to 7 days).

Our model is based on a simple method of energy balance combined with heat conduction. The model data are based on the building design data. The model is validated using real data from the same building.

We tested the model based on real conditions over a 14-day period with and without ventilation air entering the building. The model is accurate within +/-1°C for 7 days prediction for cases where the ventilation system is off.

For simulation with the ventilation system in normal operation, the model predicts the change in room temperature within an accuracy of +/-1°C for 2 days. Additional accuracy for longer periods may require input of the ventilation air temperature. This should be checked in future work.

Future work should provide qualitative measures for the error between the measured and the predicted room temperature. This will help our understanding of the prediction accuracy.

We will extend the model to predict the effect of solar influx. Solar radiation is likely to affect the temperature of the walls and may result in a different thermal loss from the building.

Future models should verify the dynamic behavior of the building. The simplistic approach taken in this work is a possible oversimplification, and can lead to serious errors if not closely monitored.

References

S. Bredesen. *Energy utilization and storage in buildings*, University of South-Eastern Norway, Master's thesis, 2019.

I. Kildsgaard, A. Jarnehammar, A. Widheden, and M. Wall. Energy and Environmental Performance of Multi-Story Apartment Buildings Built in Timber Construction Using Passive House Principles. *Buildings*, 3(1):258-277, 2013. doi: 10.3390/buildings3010258.

B. Lie. Surrogate and Hybrid Models for Control. In Proceedings *SIMS 2019, Västerås, Sweden, 13-16 August, 2019. Linköping Electronic Conference Proceedings Series*, 170: 1-8, 2019. doi:10.3384/ecp201701.

A.M. Madni, C.C. Madni, and S.D. Lucero. Leveraging Digital Twin Technology in Model-Based Systems Engineering. *Systems*, 7(1):7, 2019. doi:10.3390/systems7010007.

G. Makaka. A Pedestrian Approach to Indoor Temperature Distribution Prediction of a Passive Solar Energy Efficiency House. *Journal of Renewable Energy, 1-9*, 2015. doi:10.1155/2015/128496.

N. Mendes, H.C. Oliveira, and XdA. Humberto. Building Thermal Performance Analysis by Using Matlab/Simulink. In Proceedings *7th International IBPSA Conference*, Rio de Janeiro, Brazil, 2001.

NS-EN ISO 52000-1. *Energy performance of buildings – Overarching EPB assessment – Part 1: General framework and procedures.* Standard Norge, Lysaker, Norway. NS-EN ISN 52000-1:2017.

NS3701. *Criteria for passive houses and low energy buildings Non-residential buildings*. Standard Norge, Lysaker, Norway. NS 3701:2012.

P. Rohdin, A. Molin, and B. Moshfegh. Experiences from nine passive houses in Sweden – Indoor thermal environment and energy use. *Building and Environment*, 71:176-185, 2014. doi:10.1016/j.buildenv.2013.09.017.

SIMIEN. *SW for simulation of energy consumption in buildings*. 2020. www.programbyggerne.no. Accessed 5-May-2020.

Extended model for control of thermal energy in buildings

Ashish Bhattarai, Zahir Barahmand, Sina Orangi, Bernt Lie

University of South-Eastern Norway, Porsgrunn, Norway, `Bernt.Lie@usn.no`

Abstract

The paper extends a previous model of a heated water system with stratification, with an external floor heating circulation loop and a detailed model of heat transfer from water pipe to heated room. A minor error in the previous stratification model is corrected. The floor heating loop is posed as a cross-current heat exchanger, and a simple approximation of time delay for heat advection is suggested. For heat flow to the room, infinitely fast heating of an aluminum plate is suggested, with slower heat transfer to chip-board, through fiberboard and parquet to the room. The room is heated by a combination of convection and radiation. The results show that the inclusion of the heat circulation loop shows that this has an important effect on the dynamics of the system, and that this loop should be taken into account if parameters are tuned to fit the model to experimental data.

Keywords: floor heating, energy in buildings, energy storage, hot water tank model, water distribution system, stratified flow model

1 Introduction

Floor heating has been used for thousands of years, and essentially consists of a heat generation system and a heat distribution system with heat transfer through the floor. Modern studies of low energy buildings focus on taking advantage of water with low thermal value (lukewarm, 30–35°C), which necessitates reducing heat transfer coefficients in the system. Modern control systems allow for reducing the temperature when (part of) the building is unused, but require heating system with low heat capacity to be efficient. This implies using above-floor systems, i.e., inserting pipes in the underlayment between the subfloor (e.g., chipboard) and the floor covering (parquet, etc.).

(Lie et al., 2014) discussed the use of solar heating assisted by electric heating for floor heating, and studied the use of Model Predictive Control (MPC), and (Lie, 2015) discussed a minor improvement of the heater model. In (Johansen et al., 2019), an improved model of an electric heater was considered, and compared with experimental data. Specifically, a model of stratification due to (Viskanta et al., 1977) was used, see also (Xu et al., 2014). (Lago et al., 2019) discusses a similar model, and a possible smooth description of the buoyancy conductivity. A more complex model of stratification is given in (Vrettos, 2016), with a two stage diffusion predictor and buoyancy corrector step. A system with some details of floor heating
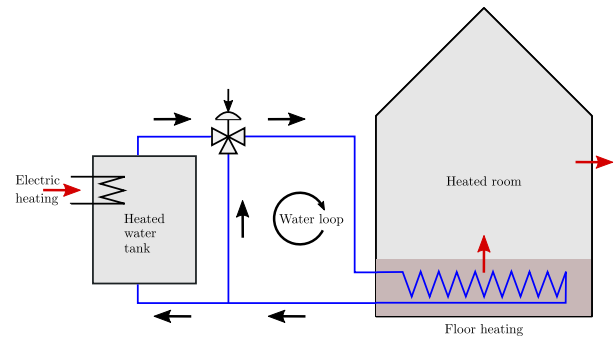
Figure 1. Floor heating system.

is discussed in (Ho et al., 1995).

In this paper, the heated water tank model as in (Johansen et al., 2019) is corrected for missing constants, and is extended with a water distribution system for floor heating. The system is extended to include a circulation loop for floor heating — with more details about the floor layers than in (Lie et al., 2014), while excluding the solar heating. The effect of the circulation loop is examined; this circulation loop was not included in the model fitting of (Johansen et al., 2019). The paper is organized as follows. In Section 2, an overview is given of the system. In Section 3, the dynamic model of the system with heater, water circulation, and floor layers + room is developed. In Section 4, some simulations results are given, while conclusions are drawn in Section 5.

2 System overview

2.1 Floor heating

Consider a floor heating system for a building, Fig. 1.

The system consists of an electrically heated, stratified water tank which supplies heated water to a water loop passing through water pipes embedded in the floor. The heated floor then provides heating to the room above to compensate for heat loss to the surroundings. Both floor temperature and air temperature in the heated room is of importance for inhabitant comfort. Typically, a floor temperature of ca. 22°C and an air temperature of ca. 20°C is deemed optimal when in use.

The heated water tank is influenced by external signals in the form of the loop volumetric water flow rate $\dot{V}_\ell$, the split range valve signal u_v which determines how much water goes through the heated tank, the ambient temperature of the heated tank, T_a^t, and the fraction of full electric power u_P that is used to heat the tank. In the model of (Jo-

hansen et al., 2019), the inlet temperature T_i at the bottom of the tank was also an external signal; in this work, T_i is a signal that comes from the heated floor subsystem, and is thus not a free, external input.

The water from the heated tank flows into the water loop at temperature T_i^{ℓ} (the temperature after the split range valve of the heated tank), and passes through a lengthy pipe embedded in the floor. The floor pipe is essentially a heat exchanger for transfer of heat to the floor. From the floor, heat is transferred to the room by convection and radiation. Finally, the room experiences a heat loss to the surroundings which is at external ambient temperature T_a^r.

2.2 Heated water tank: modification

The stratified heated water tank with discretization is given in detail in Johansen et al. (2019). In (Johansen et al., 2019), buoyant turbulent mixing flux $\dot{Q}_{db}''$ was correctly given by

$$\frac{\partial \dot{Q}_{db}''}{\partial z} = -k_b \frac{\partial^2 T}{\partial z^2}.$$

Here, z is vertical position, k_b is a buoyancy conductivity, and T is the temperature distribution over z . However, here we correct the expression for k_b, which should be

$$k_b = \begin{cases} \rho \hat{c}_p \cdot \kappa^2 d^2 \sqrt{g \alpha_p \left| \frac{\partial T}{\partial z} \right|}, & \frac{\partial T}{\partial z} < 0 \\ 0, & \frac{\partial T}{\partial z} \geq 0, \end{cases}$$

where ρ is density, $\hat{c}_p$ is specific heat capacity, κ is the von Kármán constant, d is the tank diameter, g is acceleration of gravity, and α_p is the thermal expansion coefficient at constant pressure.

The essential difference is that in the previous publication, $\rho \hat{c}_p$ was replaced by an ad hoc tuning parameter c_b which was suggested to have a value near unity.

2.3 Transport of water in pipes

If we assume that the pipes transporting water to and from the heated floor are perfectly insulated, the temperature for these stretches are given by the advection equation:

$$\frac{\partial T}{\partial t} = -v \frac{\partial T}{\partial x}$$

where

$$v = \frac{\dot{V}_{\ell}}{A_p}$$

with $\dot{V}_{\ell}$ the volumetric flow rate in the water loop through the floor heating system, and A_p the cross sectional area of the pipe. The advection model has the well known solution

$$T(t, x = L) = T\left(t - \frac{L}{v}, x = 0\right)$$

where L is the length of the pipe. The Laplace transform of this solution is

$$T(s, x = L) = \exp\left(-\frac{L}{v} s\right) T(s, x = 0).$$
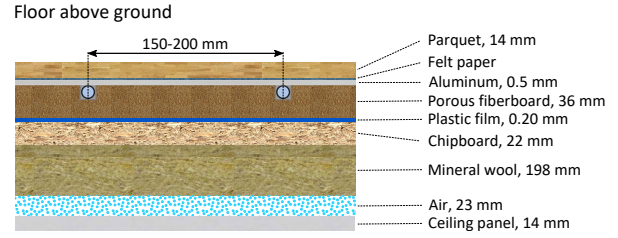
Figure 2. Structure of floor heating (not to scale), freely after https://www.uponor.no/vvs/produkter/gulvvarme/calma-trinnlydplater. Water flows in water pipes, which typically are separated by 150–200 mm.

This time delay can be expressed directly in some modeling languages. In Modelica, the syntax would be[1]:

```
TxL = delay(Tx0, L/v)
```

Other languages, such as Julia, have special solvers for delay differential equations.

Introducing $\tau \equiv \frac{L}{v}$ as the time delay, we can approximate the solution by N compartments. With partial delay τ_i such that $\sum_{i=1}^N \tau_i = \tau$, a possible Padé approximation is

$$\frac{T(s; x = L)}{T(s; x = 0)} = \frac{1}{\exp(\tau_N s)} \prod_{i=1}^{N-1} \frac{\exp\left(-\frac{\tau_i}{2} s\right)}{\exp\left(\frac{\tau_i}{2} s\right)} \approx \frac{1}{1 + \tau_N s} \prod_{i=1}^{N-1} \frac{1 - \frac{\tau_i}{2} s}{1 + \frac{\tau_i}{2} s}$$

with realization

$$\frac{dx_1}{dt} = \frac{1}{\tau_1}(T_{x=0} - x_1)$$

$$\frac{dx_1}{dt} + \frac{dx_2}{dt} = \frac{2}{\tau_2}(x_1 - x_2)$$

$$\vdots$$

$$\frac{dx_{N-1}}{dt} + \frac{dx_N}{dt} = \frac{2}{\tau_N}(x_{N-1} - x_N)$$

where $T_{x=L} = x_N$. An advection model/time delay has non-minimum phase characteristics (Åstrøm and Murray, 2008), and the all-pass terms $\frac{1 - \frac{\tau_i}{2} s}{1 + \frac{\tau_i}{2} s}$ ensures that the non-minimum phase characteristic is retained. At the same time, the lag term $\frac{1}{1 + \tau_N s}$ ensures that the time derivative of the input signal $T_{x=0}$ can be avoided. This advection model approximation constitutes a DAE, which can easily be changed into an ODE. However, the resulting ODE becomes more complicated, and if we use a DAE solver, such a reformulation is not needed.

2.4 Heat transfer from water to floor

2.4.1 Structure of heated floor

The structure of the floor layers is depicted in Fig. 2.

Water pipes of PEX (cross-linked polyethylene) are put in grooves in the *underlayment*, typically in flexible, porous fiberboard which also serves to dampen the sound

---

[1] https://www.openmodelica.org/forum/default-topic/1907-how-to-make-a-time-delay,-w-t-r

of steps. To secure good heat transfer to the *floor finish* (e.g., parquet), the pipes are put in thermal contact with an aluminum plate which has high thermal conductivity, and thus spreads the heat evenly in the aluminum plate. The water pipes are either surrounded (50-75%) by Ω-plates or clips, or they are surrounded by a plaster with good thermal conductivity such as *anhydrite* — this is to secure the best possible heat transfer area to the aluminum plate.

To achieve even temperature in the aluminum plate and thus in the floor finish (e.g., parquet), pipes of outer diameter 12–20 mm and 2 mm wall thickness of PEX are typically put at a distance of 15–20 cm along the entire floor. Between the floor finish and the aluminum plate is felt paper, or similar, to enable some sliding between the floor finish and the underlayment and thus reduce sounds of movement.

The floor layer structure in Fig. 2 is common when there is a room below the floor. We will neglect heat transfer resistance in the very thin plastic film and felt paper. For simplicity, we will also assume that the insulation (mineral wool) is perfect, and that there is no heat leakage to the room below.

2.4.2 Heat transfer from water to aluminum

To simplify the model, we assume a fast heat transfer between water and aluminum, and due to aluminum's high heat conduction, the aluminum temperature is taken to be homogeneous. Heat transfer can then be modeled as in a steady state cross current heat exchanger. The water temperature is at the tube side (pipe), with temperature $T_p(x)$ varying along the pipe length, while aluminum is the shell side with temperature T_{al} being independent of position.

Introducing the tube side water pipe dimensionless *Stanton* number N_{St}^p,

$$N_{St}^p = \frac{\mathscr{U}_x A_x}{\dot{m}_p \hat{c}_{p,w}},$$

with heat transfer coefficient $\mathscr{U}_x$, contact surface A_x in heat exchanger, tube side (water) mass flow rate $\dot{m}_p$, and tube side heat capacity $\hat{c}_{p,w}$, and assuming N_{St}^p is independent of position x, we find the heat transfer from tube side water *pipe* to shell side *aluminum* $\dot{Q}_{p2al}$ to be

$$\dot{Q}_{p2al} = \dot{m}_p \hat{c}_{p,w} \left[1 - \exp\left(-N_{St}^p\right)\right] \left(T_i^p - T_{al}\right).$$

Here, T_i^p is the water pipe inlet temperature to the "heat exchanger", while T_{al} is the exit temperature of the shell side = aluminum temperature.

The aluminum (shell side) temperature is given by the energy balance for the shell side (aluminum). Since the heat conduction in aluminum is fast, we will assume steady state for aluminum, i.e.,

$$0 = \frac{dU_{al}}{dt} = \dot{Q}_{p2al} - \dot{Q}_{al2x},$$

where U_{al} is the internal energy of the shell side (aluminum), $\dot{Q}_{al2x}$ is the combined heat flow to the fiberboard and the parquet. We will return to an expression for $\dot{Q}_{al2x}$.

We also need the effluent water (tube side) temperature T_e^p, which is the temperature of the water returning to the heated tank:

$$T_e^p = T_{al} + \exp\left(-N_{St}^p\right)\left(T_i^p - T_{al}\right). \tag{1}$$

In the expression for the Stanton number,

$$\dot{m}_p = \dot{V}_\ell \rho_w$$

$$\frac{1}{\mathscr{U}_x} = \frac{1}{h_w} + \frac{\frac{d_p}{2} \ln\left(1 + \frac{2\delta_p}{d_p}\right)}{k_p}$$

where h_w is the heat transfer coefficient for water to solid in the pipe, d_p is the inner diameter of the pipe, δ_p is the thickness of the plastic water pipe wall, and k_p is thermal conductivity of the pipe material (plastic) (Lydersen, 1979).

Likewise, the contact area A_x is, say, 75% of the plastic pipe surface, i.e.,

$$A_x = 0.75 \cdot \pi \left(d_p + 2\delta_p\right) L_f$$

where d_p is the inner diameter of the pipe, while L_f is the length of the pipe in contact with the shell side (aluminum) in the *floor*.

2.4.3 Heat transfer to plates

Steady energy balance for the aluminum plate leads to

$$\dot{Q}_{p2al} = \dot{Q}_{al2x} = \dot{Q}_{al2pq} + \dot{Q}_{al2fb}$$

where $\dot{Q}_{al2pq}$ is the heat transfer from aluminum at temperature T_{al} to parquet, while $\dot{Q}_{al2fb}$ is the heat transfer from aluminum to fiberboard. In addition to this, we will also need the heat transfer $\dot{Q}_{fb2cb}$ from fiberboard to chipboard.

The dynamics of parquet, fiberboard, and chipboard will be considerably slower than that of aluminum, so we include a dynamic model of these plates. In a simplified description, we assume a thin boundary layer near aluminum for parquet and fiberboard, and a thin boundary layer near fiberboard for chipboard — for the rest of the board volumes, we assume homogeneous temperature T. This leads to a simplified model

$$\rho A \hat{c}_p \delta \frac{dT}{dt} = A \frac{k}{\delta}\left(T_0 - T\right) - \dot{Q}_{x=\delta}.$$

Applying this model to all three boards, we associate quantities as in Table 1. In Table 1, $\dot{Q}_{pq2r}$ is the heat flow from the parquet to the room — which consists of convective heat flow and radiative heat flow. $\dot{Q}_{fb2cb}$ is the heat flow from the fiberboard to the chipboard. Finally, the heat flow out of the chipboard is zero, because we have assumed that the insulation material below the chipboard ("mineral wool") is perfect.

<div style="text-align:center">Table 1. Quantities and flow rates for boards.</div>

Quantity	Parquet	Fiberboard	Chipboard
T	T_{pq}	T_{fb}	T_{cb}
δ	δ_{pq}	δ_{fb}	δ_{cb}
T_0	T_{al}	T_{al}	T_{fb}
$\dot{Q}_{x=\delta}$	$\dot{Q}_{pq2r}$	$\dot{Q}_{fb2cb}$	0

2.5 Heat transfer related to room

The heat transferred into the room from the parquet floor, $\dot{Q}_{pq2r}$, consists of heat convection from the floor, $\dot{Q}_{pq2r}^c$, and net radiation, $\dot{Q}_{pq2r}^r$. Heat convection is given by floor area A_{pq} and heat flux $\dot{Q}_{pq2r}^{\prime\prime,c}$,

$$\dot{Q}_{pq2r}^{\prime\prime,c} = \mathscr{U}_{pq2r}\left(T_{pq} - T_r\right).$$

The net radiation consists of radiation out due to parquet floor temperature, and return radiation due to radiation from the ceiling. Assuming that the ceiling has the same temperature as the room air T_r, it can be shown that

$$\dot{Q}_{pq2r}^{\prime\prime,r} = \frac{1}{\frac{1}{\varepsilon_{pq}} + \frac{1}{\varepsilon_r} - 1}\sigma\left(T_{pq}^4 - T_r^4\right)$$

where σ is Stefan-Boltzmann's constant, and ε_{pq} and ε_r are the emissivities from the parquet floor and the room (ceiling), respectively. In this radiation expression, absolute temperature must be used. The expression is based on radiation between two parallel planes. In summary,

$$\dot{Q}_{pq2r} = A_{pq}\left(\dot{Q}_{pq2r}^{\prime\prime,c} + \dot{Q}_{pq2r}^{\prime\prime,r}\right).$$

There is also a convective heat loss to the surroundings, $\dot{Q}_{r2a}$, given by

$$\dot{Q}_{r2a} = A_r\dot{Q}_{r2a}^{\prime\prime}$$

where A_r is the net surface between the room and the ambient of the room, while

$$\dot{Q}_{r2a}^{\prime\prime} = \mathscr{U}_{r2a}\left(T_r - T_a^r\right).$$

3 Dynamic model

Since the density of water and air will be assumed constant, we essentially need the energy balance. The model can be summarized as follows.

3.1 Heated tank

The model from (Johansen et al., 2019) has been corrected in a project in a course at USN[2], as described in Section 2.2. For model details, see (Johansen et al., 2019).

---

[2]University of South-Eastern Norway: Course FM1015 Modelling of Dynamic Systems, group project Fall of 2019.

3.2 Floor heating/heat exchanger

If we neglect the time delay from the heated tank to the inlet to the floor coil pipes, we have $T_i^p(t) = T_i^\ell(t)$. If we instead include the time delay, we have $T_i^p(t) = T_i^\ell(t - \tau_{t2f})$ where the time delay from the heater to the floor is $\tau_{t2f} = \frac{L_{t2f}}{v} = \frac{L_{t2f}}{\dot{V}_\ell/A_p}$ with length L_{t2f} of the water pipe from the heated tank to the floor coil inlet, volumetric flow rate $\dot{V}_\ell$ in the floor heating pipes, and cross sectional area A_p of the pipe, i.e., $A_p = \pi r_p^2 = \pi d_p^2/4$.

With aluminum plate temperature T_{al}, the effluent temperature T_e^p of the water after the floor has been heated is then

$$T_e^p = T_{al} + \exp\left(-N_{St}^p\right)\left(T_i^p - T_{al}\right),$$

where the tube (pipe) side dimensionless *Stanton* number N_{St}^p is

$$N_{St}^p = \frac{(\mathscr{U}A)_x}{\dot{m}_p\hat{c}_{p,w}},$$

with overall heat transfer coefficient $\mathscr{U}_x$, contact surface A_x in heat exchanger, water mass flow rate in the pipes $\dot{m}_p$, and water heat capacity $\hat{c}_{p,w}$, and assuming N_{St}^p is independent of position x. Here, the overall heat transfer coefficient is given by

$$\frac{1}{\mathscr{U}_x} = \frac{1}{h_w} + \frac{\frac{d_p}{2}\ln\left(1 + \frac{2\delta_p}{d_p}\right)}{k_p}$$

where h_w is the heat transfer coefficient from water to pipe wall, k_p is the conductivity of the pipe wall (plastic), d_p is the inner pipe diameter, while δ_p is the pipe thickness. The contact surface A_x is assumed to be 75% of the external surface of the pipes, i.e.,

$$A_x = 0.75\pi\left(d_p + 2\delta_d\right)L_f.$$

The heat transferred from water pipe (tube side) to aluminum (shell side) is then $\dot{Q}_{p2al}$ given by

$$\dot{Q}_{p2al} = \dot{m}_p\hat{c}_{p,w}\left[1 - \exp\left(-N_{St}^p\right)\right]\left(T_i^p - T_{al}\right).$$

Here, T_i^p is the water pipe inlet temperature to the "heat exchanger", while T_{al} is the exit temperature of the shell side; since the shell side is assumed to have homogeneous, T_{al} is the aluminum temperature.

The water temperature that enters the return loop to the heated tank is $T_e^p(t)$ if we neglect time delay, and $T_e^p(t - \tau_f)$ if we include the time delay of the water flowing through the floor coils, $\tau_f = \frac{L_f}{v} = \frac{L_f}{\dot{V}_\ell/A_p}$ (assuming the same pipe cross sectional area everywhere).

Assuming steady energy balance for the aluminum plate gives

$$0 = \dot{Q}_{p2al} - \dot{Q}_{al2x},$$

where $\dot{Q}_{p2al}$ is given above, while $\dot{Q}_{al2x}$ is given by

$$\dot{Q}_{al2x} = \dot{Q}_{al2pq} + \dot{Q}_{al2fb}$$

with

$$\dot{Q}_{\text{al2pq}} = A_{\text{r}} \frac{k_{\text{pq}}}{\delta_{\text{pq}}} \left(T_{\text{al}} - T_{\text{pq}} \right)$$

and

$$\dot{Q}_{\text{al2fb}} = A_{\text{r}} \frac{k_{\text{fb}}}{\delta_{\text{fb}}} \left(T_{\text{al}} - T_{\text{fb}} \right)$$

with A_{r} being the area of the room floor.

For the heat flow from fiberboard to chipboard, we have

$$\dot{Q}_{\text{fb2cb}} = A_{\text{r}} \frac{k_{\text{cb}}}{\delta_{\text{cb}}} \left(T_{\text{fb}} - T_{\text{cb}} \right).$$

The effluent water from floor heating at temperature T_{e}^{p} is returned to the heated tank through a pipe of length L_{f2t}. If the time delay is neglected, we have $T_{\text{i}}(t) = T_{\text{e}}^{\text{p}}(t)$. If we include the time delay, the relation is $T_{\text{i}}(t) = T_{\text{e}}^{\text{p}}\left(t - \tau_{\text{f2p}} \right)$ where $\tau_{\text{f2p}} = \frac{L_{\text{f2p}}}{\dot{V}_{\ell}/A_{\text{p}}}$.

3.3   Board models

The models for the parquet, the fiberboard and the chipboard can be summarized as follows:

$$\rho_{\text{pq}} A_{\text{r}} \hat{c}_{p,\text{pq}} \delta_{\text{pq}} \frac{dT_{\text{pq}}}{dt} = \dot{Q}_{\text{al2pq}} - \dot{Q}_{\text{pq2r}}$$

$$\rho_{\text{fb}} A_{\text{r}} \hat{c}_{p,\text{fb}} \delta_{\text{fb}} \frac{dT_{\text{fb}}}{dt} = \dot{Q}_{\text{al2fb}} - \dot{Q}_{\text{fb2cb}}$$

$$\rho_{\text{cb}} A_{\text{r}} \hat{c}_{p,\text{cb}} \delta_{\text{cb}} \frac{dT_{\text{cb}}}{dt} = \dot{Q}_{\text{fb2cb}}.$$

Here,

$$\dot{Q}_{\text{pq2r}} = \dot{Q}_{\text{pq2r}}^{\text{c}} + \dot{Q}_{\text{pq2r}}^{\text{r}}$$

$$\dot{Q}_{\text{pq2r}}^{\text{c}} = A_{\text{r}} \mathscr{U}_{\text{pq2r}} \left(T_{\text{pq}} - T_{\text{r}} \right)$$

$$\dot{Q}_{\text{pq2r}}^{\text{r}} = A_{\text{r}} \frac{1}{\frac{1}{\varepsilon_{\text{pq}}} + \frac{1}{\varepsilon_{\text{r}}} - 1} \sigma \left(T_{\text{pq}}^{4} - T_{\text{r}}^{4} \right).$$

3.4   Room model

With a simplistic room model, we only consider air mass m_{r} with no ventilation. Then

$$\frac{dU_{\text{r}}}{dt} = \dot{Q}_{\text{pq2r}} - \dot{Q}_{\text{r2a}}$$

where U_{r} is the internal energy of the room air, for simplicity, $dU_{\text{r}} \approx m_{\text{r}} \hat{c}_{v,\text{a}} dT_{\text{r}}$, where $\hat{c}_{v,\text{a}} = \hat{c}_{p,\text{a}} - \frac{R}{M_{\text{a}}}$ is the specific heat capacity of air at constant volume, $\hat{c}_{p,\text{a}}$ is the specific heat capacity at constant pressure, with gas constant R and molar mass of air M_{a}. Here,

$$\dot{Q}_{\text{r2a}} = A_{\text{r}}^{\text{s}} \mathscr{U}_{\text{r}} \left(T_{\text{r}} - T_{\text{a}}^{\text{r}} \right)$$

where A_{r}^{s} is the surface area of the room against the ambient temperature T_{a}^{r}, $\mathscr{U}_{\text{r}}$ is the overall heat transfer coefficient from room air temperature T_{r} to ambient temperature.

In a more realistic room model, we would also take into account stored energy in furniture, walls, etc., multiple rooms with transport between the rooms, and radiation from the sun into the room.

3.5   Model parameters

Model parameters for the heated tank are given in (Johansen et al., 2019), while model parameters for the floor/heated room are given in Table 2.

Let us briefly discuss the time delays of water flow in the system. With $\dot{V}_{\ell} \in [1,13]\,\text{L/min}$ and $A_{\text{p}} = \pi d_{\text{p}}^{2}/4$ we find time delays:

$$\tau_{\text{t2f}} = \frac{L_{\text{t2f}}}{\dot{V}_{\ell}/A_{\text{p}}} = \frac{20 \cdot 10^{1} \pi \frac{\left(12 \cdot 10^{-2} \right)^{2}}{4}}{\dot{V}_{\ell}} = \frac{2.26}{\dot{V}_{\ell}} = [0.17, 2.3]\,\text{min}$$

$$\tau_{\text{f}} = \frac{L_{\text{f}}}{\dot{V}_{\ell}/A_{\text{p}}} = \frac{250}{\dot{V}_{\ell} \pi \frac{\left(12 \cdot 10^{-2} \right)}{4}} = [2.2, 28.3]\,\text{min}$$

$$\tau_{\text{f2t}} = \frac{L_{\text{f2t}}}{\dot{V}_{\ell}/A_{\text{p}}} = \tau_{\text{t2f}}.$$

This means that for low flow rates, the time delay main be up to 30 min through the floor pipe. The time delay in the transport pipes is small, though. The characteristic time constants of the system are in the same order of magnitude, (Johansen et al., 2019). Thus, the time delay should be considered. For simplicity, a simple approximation of the time delay is to put it between the outlet of the floor pipe and the heated tank. We use the approximate description

$$\frac{dx_{1}}{dt} = \frac{1}{\tau_{1}} \left(T_{\text{e}}^{\text{p}} - x_{1} \right)$$

$$\frac{dx_{1}}{dt} + \frac{dx_{2}}{dt} = \frac{2}{\tau_{2}} \left(x_{1} - x_{2} \right)$$

$$\vdots$$

$$\frac{dx_{N-1}}{dt} + \frac{dx_{N}}{dt} = \frac{2}{\tau_{N}} \left(x_{N-1} - x_{N} \right)$$

with $T_{\text{i}} = x_{N}$ and $\tau = \sum_{i=1}^{N} \tau_{i} = \frac{(L_{\text{t2f}} + L_{\text{f}} + L_{\text{f2t}}) A_{\text{p}}}{\dot{V}_{\ell}}$. For simplicity, we set $\tau_{i} = \frac{\tau}{N}$.

4   Simulation results

The model of the heated tank with 20 slices and specified input temperature T_{i} is simulated, and compared to a model of the combined heated tank and floor heating with circulating water providing T_{i}, using the model parameters in (Johansen et al., 2019) and those in Table 2, with $N = 3$ "volumes" in the time delay approximation.

Key inputs to the systems are given in Fig. 3. Observe that inlet temperature to the heated tank, T_{i}, is only used when the heated tank is simulated as an independent system. It has been assumed that the ambient temperature to the heated tank is $T_{\text{a}}^{\text{t}} = 15\,°\text{C}$, while that the outdoor temperature is $T_{\text{a}}^{\text{r}} = 5\,°\text{C}$; both are assumed to be constant over time. Initial values of all states (temperatures) are set to $25\,°\text{C}$.

Figure 4 compares the temperature distribution in the heated tank for the cases (a) that the heated tank is an independent system with specified T_{i}, and (b) that the heated

Table 2. Model parameters for room with floor heating.

Parameter	Value	Comment
d_p	$12\,\mathrm{mm}$	Inner diameter of floor pipe
δ_p	$2\,\mathrm{mm}$	Thickness of floor pipe wall
k_p	$0.5\,\mathrm{W/m\,K}$	Thermal conductivity of floor pipe material (PEX)
L_f	$250\,\mathrm{m}$	Length of pipe in floor heat exchanger
L_t2f	$20\,\mathrm{m}$	Length of pipe from tank to floor coils
L_f2t	$20\,\mathrm{m}$	Length of return pipe from floor coils to tank
A_x	$0.75\pi\left(d_\mathrm{p}+2\delta_\mathrm{p}\right)L_\mathrm{f}$	Heat transfer area from water to aluminum plate
h_w	$6500\,\mathrm{W/m^2K}$	Heat transfer from water to pipe wall
$\mathscr{U}_\mathrm{x}$	$1/\left(\frac{1}{h_\mathrm{w}}+\frac{\frac{d_\mathrm{p}}{2}\ln\left(1+\frac{2\delta_\mathrm{p}}{d_\mathrm{p}}\right)}{k_\mathrm{p}}\right)$	Overall heat transfer coefficient, water to aluminum
A_r	$50\,\mathrm{m^2}$	Area of floor with floor heating
h_r	$2.5\,\mathrm{m}$	Height of room
V_r	$h_\mathrm{r}A_\mathrm{r}$	Volume of room
ρ_a	$1.225\,\mathrm{kg/m^3}$	Density of air
$\hat{c}_{p,\mathrm{a}}$	$1\,\mathrm{kJ/kg\,K}$	Specific heat capacity at constant pressure, air
R	$8.31\,\mathrm{kJ/kmol\,K}$	Gas constant
M_a	$28.97\,\mathrm{kg/kmol}$	Molar mass of air
$\hat{c}_{v,\mathrm{a}}$	$\hat{c}_{p,\mathrm{a}}-\frac{R}{M_\mathrm{a}}$	Specific heat capacity at constant volume, air
m_r	$\rho_\mathrm{a}V_\mathrm{r}$	Mass of air in room
δ_pq	$14\,\mathrm{mm}$	Thickness of parquet
δ_fb	$36\,\mathrm{mm}$	Thickness of fiberboard
δ_cb	$22\,\mathrm{mm}$	Thickness of chipboard
ρ_pq	$750\,\mathrm{kg/m^3}$	Density of parquet material
ρ_fb	$230\,\mathrm{kg/m^3}$	Density of fibreboard material
ρ_cb	$700\,\mathrm{kg/m^3}$	Density of chipboard material
$\hat{c}_{p,\mathrm{pq}}$	$2\,\mathrm{kJ/kg\,K}$	Specific heat capacity of parquet material
$\hat{c}_{p,\mathrm{fb}}$	$1.4\,\mathrm{kJ/kg\,K}$	Specific heat capacity of fibreboard material
$\hat{c}_{p,\mathrm{cb}}$	$1.8\,\mathrm{kJ/kg\,K}$	Specific heat capacity of chipboard matierial
k_pq	$0.17\,\mathrm{W/m\,K}$	Thermal conductivity of parquet material
k_fb	$0.049\,\mathrm{W/m\,K}$	Thermal conductivity of fibreboard material
k_cb	$0.15\,\mathrm{W/m\,K}$	Thermal conductivity of chipboard material
σ	$5.6494\cdot10^{-8}\,\frac{\mathrm{W}}{\mathrm{K^4\,m^2}}$	Stefan-Boltmann's constant
ε_pq	$0.9\,-$	Emissivity of parquet
ε_r	$0.96\,-$	Emissivity of ceiling (room)
A_r^s	$A_\mathrm{r}+4h_\mathrm{r}\sqrt{A_\mathrm{r}}$	Surface area of room
$\frac{k_\mathrm{r}}{\delta_\mathrm{r}}$	$0.15\,\mathrm{W/m^2K}$	Heat transfer coefficient through wall
$\mathscr{U}_\mathrm{r}$	$1/\left(\frac{1}{h_\mathrm{a}}+\frac{\delta_\mathrm{r}}{k_\mathrm{r}}+\frac{1}{h_\mathrm{a}}\right)$	Overall heat transfer coefficient, room to ambient
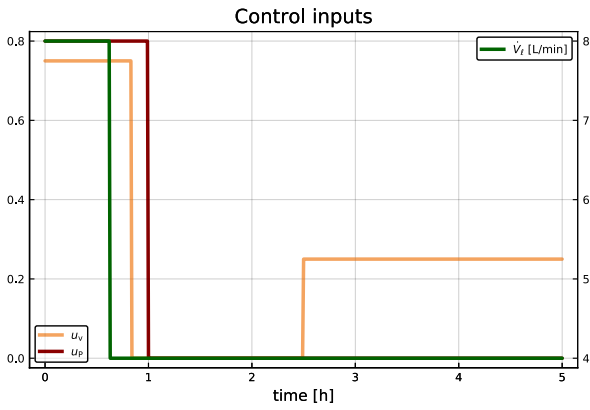
Figure 3. Control inputs u_P (power fraction to heated tank), u_v (water flow valve opening), and $\dot{V}_\ell$ (volumetric flow rate in heating loop).
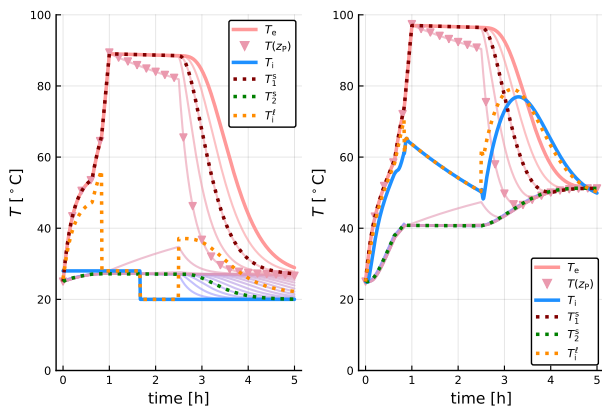
Figure 4. Temperature distribution (T_e: effluent temperture from heated tank; $T(z_P)$: temperature at heating element; T_i: influent temperature to heated tank; T_j^s temperature at heated tank sensors; T_i^ℓ inlet temperature to floor heating loop) in heated tank for cases (a) the heated tank is an independent system with specified tank inlet temperature T_i, and (b) the heated tank and the floor heating system are connected via a water loop where T_i is the return water temperature from the floor heating.

Figure 5. Temperature evolution in various compartments in the floor heating system (T_j: temperature in compartment j — al: aluminum plate, pq: parquet, r: room, fb: fiber board, cb: chip board). Observe that all compartments start at 25 °C.
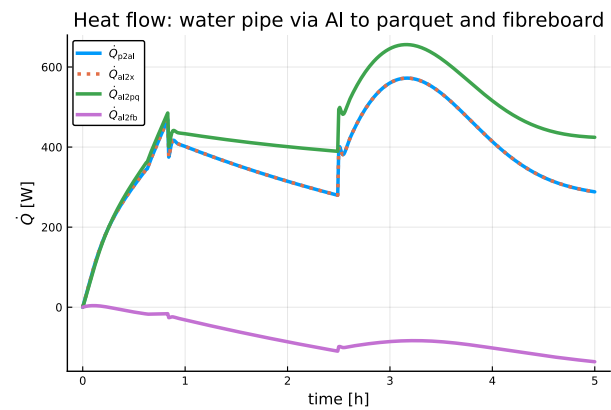
Figure 6. Heat flows from floor water pipe through aluminum to parquet and fiberboard ($\dot{Q}_{i2j}$: heat flow rate from compartment i to j, where i, j are p: water pipe, al: aluminum plate, pq: parquet, fb: fiber board). Observe that for the case studied here, the flow from the aluminum plate to the fibreboard is negative.
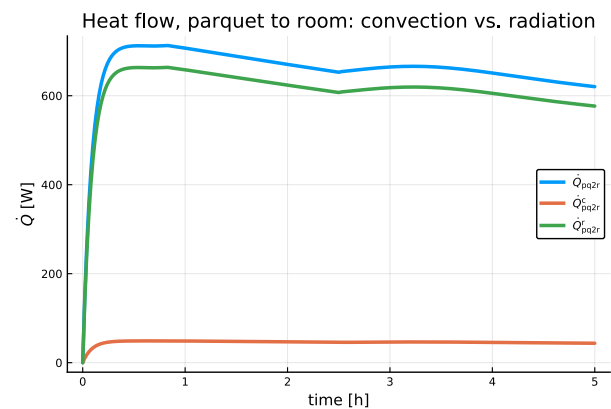
Figure 7. Heat flows from parquet to room ($\dot{Q}_{i2j}^k$: heat flow rate from compartment i to compartment j, where pq: parquet, r: room, and k indicates heat flow type, c: convection, r: radiation).

tank and the floor heating system are connected such that T_i is caused by a return of water from the floor.

Figure 5 shows the temperatures in the floor heating part: T_{al}, T_{pq}, T_r, T_{fb}, T_{cb}.

Using a DAE solver (OpenModelica), we immediately also find other quantities such as heat flows. As an example, Fig. 6 shows the heat flow from water pipe to aluminum, $\dot{Q}_{al2p}$, as well as heat flow out of the aluminum plate, $\dot{Q}_{al2x}$. Furthermore, the figure shows the separate heat flow from aluminum to parquet, $\dot{Q}_{al2pq}$, and from aluminum to fiberboard, $\dot{Q}_{al2fb}$.

Because steady state is assumed for the aluminum plate, $\dot{Q}_{al2x} \equiv \dot{Q}_{p2al}$. Furthermore, $\dot{Q}_{al2x} = \dot{Q}_{al2pq} + \dot{Q}_{al2fb}$. Observe that with the given initial temperatures of the fiberboard, $\dot{Q}_{al2fb} < 0$ in these operating conditions.

Figure 7 shows the total heat flow from the parquet to the room, and the heat flow due to convection vs. radiation.

An interesting observation is that heat flow by radiation dominates over heat flow by convection. Often, the radiation flow term is linearized and converted into a quasi convection term, with modification of the total heat transfer coefficient.

5   Conclusions

A previous model of a stratified heated water tank has been improved, and extended with a water circulation loop for heating of the floor in a room. A simple approximation of an advection model has been used in the water loop, and the heat transfer to the floor layers has been described as a cross-current heat exchanger to an aluminum plate with infinitely fast dynamics. The heat is then transported from the aluminum plate to a chip-board below the aluminum plate, and through a fiberboard plate through the parquet, and then via convection and radiation to the room. In the floor layers, the effects of a thin plastic film and a thin felt paper have been neglected, and perfect insulation has been assumed below the chipboard. The room model is overly simple in that only air has been included; in reality, air heats furniture, walls, etc., which will add to the thermal mass of the room. Still, the extension in the paper are believed to give a realistic description of the transfer of heat from a heated tank to the floor of a building.

An important result is that the water circulation loop has a considerable effect on the dynamics of the water tank temperatures. A previous paper discussed tuning of parameters for model fitting. The effect of the circulation loop indicates that the circulation loop should be taken into consideration when tuning model parameters.

The model contains a large number of parameters; (Johansen et al., 2019) and Table 2. Most of these parameters are available from the literature/experimental work at building institutes. Because of the physical nature of the model, such literature values will not be too far off from their real values. However, some of the parameters are uncertain. This is especially true with some heat transfer coefficients. Thus, in practice, it will be necessary to tune some of the parameter values based on available experimental data. Because of the physical nature of the model, the model can be expected to generalize better to other buildings than a purely empirical/data-driven model would have.

Future work will include a more detailed room/apartment model, and better scaling of the heater power compared to the heat loss in the system. It is also of interest to include more formal description of water physics and heat transfer. Model fitting will be an important part of an improved model. Finally, it is of interest to look more into how such a model can be used in a control system.

References

Karl Johan Åstrøm and Richard M. Murray. *Feedback Systems. An Introduction for Scientists and Engineers*. Princeton University Press, Princeton, NJ, 2008. ISBN 978-0-691-13576-2.

S.Y. Ho, R.E. Hayes, and R.K. Wood. Simulation of the dynamic behaviour of a hydronic floor heating systems. *Heat Recovery Systems & CHP*, 15(6):505–519, 1995.

Casper Amandus Johansen, Bernt Lie, and Nils-Olav Skeie. Models for control of thermal energy in buildings. In Erik Dahlquist, Esko Juuso, Bernt Lie, and Lars Eriksson, editors, *Proceedings of the 60th Conference on Simulation and Modelling (SIMS 59)*, number 170 in Linköping Electronic Conference Proceedings. Linköping University Electronic Press, 2019. doi:https://doi.org/10.3384/ecp20170.

Jesus Lago, Fjo De Ridder, Wiet Mazairac, and Bart De Schutter. A 1-dimensional continuous and smooth model for thermally stratified storage tanks including mixing and buoyancy. *Applied Energy*, 248:640–655, 2019.

Bernt Lie. Improved model for solar heating of buildings. In Lena Buffoni, Adrian Pop, and Bernhard Thiele, editors, *Proceedings, The 56th Conference of Simulation and Modelling*, pages 299–308, Linköping University, Sweden, October 2015. Scandinavian Simulation Society, Linköping University Electronic Press. doi:10.3384/ecp15119, pp. 299–308. October 7–9 2015.

Bernt Lie, Carlos Pfeiffer, Nils-Olav Skeie, and Hans-Georg Beyer. Models for solar heating of buildings. In Alireza Rezania Kolai, Kim Sørensen, and Mads Pagh Nielsen, editors, *Proceedings, 55th International Conference of Scandinavian Simulation Society*, pages 28–38, Aalborg University, Denmark, October 2014. Scandinavian Simulation Society, Linköping University Electronic Press. doi:www.ep.liu.se/ecp/108/ecp14108.pdf. October 21–22 2014.

Aksel L. Lydersen. *Fluid Flow and Heat Transfer*. John Wiley & Sons, Chichester, 1979.

R. Viskanta, M. Behnia, and A. Karalis. Interferometric observations of the temperature structure in water cooled or heated from above. *Advances in Water Resources*, 1(2):57–69, 1977.

Evangelos Vrettos. *Control of Residential and Commercial Loads for Power System Ancillary Services*. PhD thesis, ETH, Power Systems Laboratory, 2016.

Zhijie Xu, Ruisheng Diao, Shuai Lu, Jianming Lian, and Yu Zhang. Modeling of electric water heaters for demand response: A baseline pde model. *IEEE Transactions on Smart Grid*, 5(5):2203–2210, September 2014. doi:10.1109/TSG.2014.2317149.

Analysis of model for control of thermal energy in buildings

Sina Orangi, Zahir Barahmand, Ashish Bhattarai, Bernt Lie

University of South-Eastern Norway, Porsgrunn, Norway, `Bernt.Lie@usn.no`

Abstract

The paper considers a model for floor heating in buildings, with an electrically heated, stratified water tank, a circulation loop, a detailed description of heat transport through the floor, and a simplistic room model. The model structure is suitable for control of room temperature, and the paper discusses the use of modern simulation tools for control-relevant analysis of such models. The stratification description contains a non-differentiable buoyancy term, and two approximations are studied which circumvent this problem. The results indicate that a boundary layer approach is superior to a log-sum-exp approximation. Important basic control-relevant analysis ideas include step response (time constants), sensitivity analysis (parameter identifiability), the system zeros location (attainable performance), and Bode plots (control synthesis). The paper illustrates how such an analysis can be carried out using tools such as OpenModelica in combination with scripting language Julia.

Keywords: thermal building model, heated water tank, floor heating, model analysis, sensitivity analysis, linearization, control architecture

1 Introduction

Modern floor heating systems utilize "low quality" thermal energy in lukewarm water (30–35 °C). To reduce the overall energy consumption, it is necessary to allow for rapid temperature changes when occupants leave or arrive, hence a low heat capacity in the distribution system is desirable. Maximizing the temperature in the buildings requires low heat transfer coefficients. Building energy management systems (BEMS) are used to handle the heating and water distribution.

(Lie et al., 2014) discussed the use of solar heating assisted by electric heating for floor heating, and studied the use of Model Predictive Control (MPC). In (Johansen et al., 2019), an improved model of an electric heater was considered, and compared with experimental data. Specifically, a model of stratification due to (Viskanta et al., 1977) was introduced, see also (Xu et al., 2014). A more complex model of stratification is given in (Vrettos, 2016), with a two stage diffusion predictor and buoyancy corrector step. In (Bhattarai et al., 2020), a minor correction of the buoyancy term is introduced in reference to (Johansen et al., 2019), and the system is extended to include a circulation loop for floor heating — with more details about the floor layers than in (Lie et al., 2014), while excluding the solar heating.

Traditional heating systems for buildings often use simple temperature controllers such as thermostats. It is of interest to also consider and analyze building models from a more classical control theoretical point of view. Open loop studies typically include step response tests. Interesting questions arise related to identifiability of model parameters, where parameter sensitivity plays a key role, (Varma et al., 1999), (Jayakumar et al., 2011), (Sarmiento Ferrero et al., 2006). Model structure and input-output relations may put restrictions on attainable performance in dynamic systems. Essentially, unstable zero dynamics limits how fast feedback control loops can be made while retaining robustness. For linear models, the modes of the zero dynamics equal the system zeros, (Kwakernaak and Sivan, 1972), (Lie, 1995), and the performance limiting zeros are the right-half-plane zeros. Classical linear control design is often based on linearized approximations of models, and the Bode plot is a simple presentation of some dynamic properties for importance in control design, (Åstrøm and Murray, 2008).

The emphasis of this paper is on how to use modern computer languages to analyze the model, with the ultimate purpose of control design. We consider the languages Julia (Rackauckas and Nie, 2017) and Modelica (Fritzson, 2015), specifically OpenModelica operated via the OMJulia API (Lie et al., 2019). The focus is not on control design, but on analysis tools. The floor heating model in (Bhattarai et al., 2020) is used in the study. This model contains a $\max{()}$ function in the description of stratification, which is unfortunate when linear approximations are sought. Thus, two approximations to the max function are considered. Basic step responses are used to find the open loop response. Next, computation of output sensitivity to model parameters is illustrated, together with model linearization with location of poles and zeros as well as Bode plots. The paper is organized as follows. In Section 2, the system under study is presented, and two approaches to model approximation are presented. In Section 3, the original stratification model is compared to the two approximations. In Section 4, computer tools are used for model analysis. In Section 5, some conclusions are drawn.

2 System overview

2.1 Floor heating

Consider a floor heating system for a building, Fig. 1.
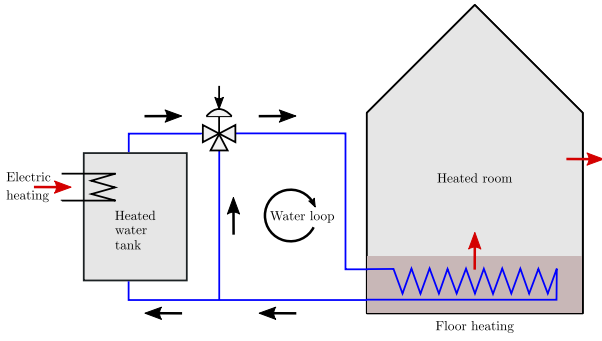
Figure 1. Floor heating system.

The system consists of an electrically heated, stratified water tank which supplies heated water to a water loop passing through water pipes embedded in the floor. The heated floor then provides heating to the room above to compensate for heat loss to the surroundings. Both floor temperature and air temperature in the heated room is of importance for inhabitant comfort. Typically, a floor temperature of ca. 22°C and an air temperature of ca. 20°C is deemed optimal when in use.

Modern buildings are not in use all the time, and it is of interest to save energy by reducing temperatures when a building is empty. To allow for quick reduction and increase in temperatures, the heat capacity of the floor should be low. On the other hand, low heat capacity leads to problems in case of power failure.

The heated water tank is influenced by external signals in the form of the loop volumetric water flow rate $\dot{V}_\ell$, the split range valve signal u_v which determines how much water goes through the heated tank, the ambient temperature of the heated tank, T_a^t, and the fraction of full electric power u_P that is used to heat the tank. In the model of (Johansen et al., 2019), the inlet temperature T_i at the bottom of the tank was also an external signal; in this work, T_i is a signal that comes from the heated floor subsystem, and is thus not a free, external input.

The water from the heated tank flows into the water loop at temperature T_i^ℓ (the temperature after the split range valve of the heated tank), and passes through a lengthy pipe embedded in the floor. The floor pipe is essentially a heat exchanger for transfer of heat to the floor. From the floor, heat is transferred to the room by convection and radiation. Finally, the room experiences a heat loss to the surroundings which is at external ambient temperature T_a^r.

2.2 Buoyancy conductivity approximations

2.2.1 Original stratification expression

The buoyancy conductivity expression k_b is

$$k_b \propto \sqrt{\max\left(-\frac{\partial T}{\partial z}, 0\right)}, \qquad (1)$$

or alternatively

$$k_b \propto \begin{cases} \sqrt{-\frac{\partial T}{\partial z}}, & \frac{\partial T}{\partial z} < 0 \\ 0, & \frac{\partial T}{\partial z} \geq 0. \end{cases} \qquad (2)$$

For implementation in Modelica, the formulation in Eq. 1 works fine, while the formulation in Eq. 2 leads to chattering between the two function branches when the gradient fluctuates around 0.

If we want to linearize the model in Modelica, things get even more complex: Modelica does not handle automatic linearization of either the expression in Eq. 2 or that in Eq. 1. Thus, for analysis purposes, it is of interest to seek approximations to the above buoyancy conductivity expressions which allow for linearizing the model by automatic linearization.

2.2.2 Log-sum-exp approximation

For simplicity, we first consider function

$$\max(-x, 0) = \begin{cases} -x, & x < 0 \\ 0, & x \geq 0. \end{cases}$$

A common approximation of this max function is the so-called "log-sum-exp" function, (Lago et al., 2019), which in a simple version can be posed as

$$\max(-x, 0) \approx \frac{1}{\mu} \log\left(\exp(-\mu x) + \exp(\mu \cdot 0)\right)$$
$$= \frac{1}{\mu} \log\left(1 + \exp(-\mu x)\right).$$

With increasing μ, this approximation becomes better and better, while still being differentiable. Thus, we can use the following approximation of the buoyancy conductivity:

$$k_b \propto \sqrt{\frac{1}{\mu} \log\left(1 + \exp\left(-\mu \frac{\partial T}{\partial z}\right)\right)}, \qquad (3)$$

where μ is tuned to give as good approximation as possible.

There is a potential problem with this approximation for our use: $\frac{1}{\mu} \log(1 + \exp(-\mu x)) > 0$ for $x > 0$. Thus, when $\frac{\partial T}{\partial z} \to x$, we will get a "buoyancy" effect when $\frac{\partial T}{\partial x} > 0$ — when there really is no buoyancy.

2.2.3 Boundary layer approximation

An alternative approach to approximation is to establish a boundary layer for $x \in [-\delta, 0]$, and use a polynomial transition from $y(-\delta)$ to $y(0)$ which simultaneously is differentiable at $x = -\delta$ and $x = 0$. For simplicity, we approximate $\sqrt{\max(-x, 0)}$ instead of $\max(-x, 0)$. It can be found that the following composed function is suitable:

$$\sqrt{\max(-x, 0)} \approx \begin{cases} \sqrt{-x}, & x < -\delta \\ \sqrt{\delta} \cdot \left(\frac{3}{2}\left(\frac{x}{\delta}\right)^3 + \frac{5}{2}\left(\frac{x}{\delta}\right)^2\right), & x \in [-\delta, 0] \\ 0, & x > 0. \end{cases} \qquad (4)$$
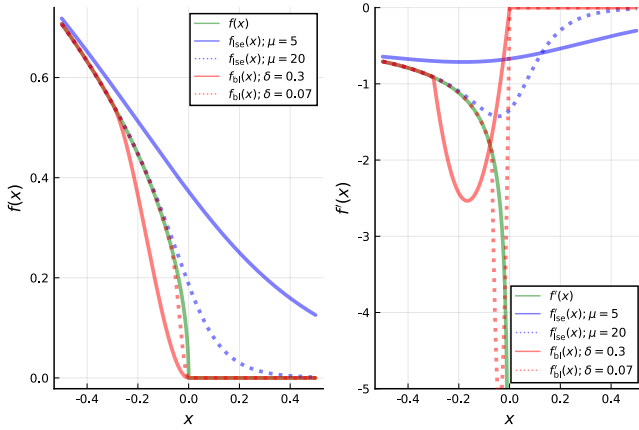
Figure 2. Left panel: comparison of $f(x) = \sqrt{\max(-x, 0)}$ (green line) with log-sum-exp approximation $f_{\text{lse}}(x) = \sqrt{\frac{\log(1+\exp(-\mu x))}{\mu}}$ in Eq. 3 (blue, solid: $\mu = 5$, dotted: $\mu = 20$), vs. boundary layer approximation $f_{\text{bl}}(x)$ in Eq. 4 (red, solid: $\delta = 0.3$, dotted: $\delta = 0.07$). Right panel: comparison of derivatives.

2.2.4 Comparison of approximations

Figure 2 illustrates the approximation of $\sqrt{\max(-x, 0)}$ with these two approaches.

Figure 2 illustrates a *potential* problem with the log-sum-exp approximation in Eq. 3 in that this approximation will give a buoyancy effect when $\frac{\partial T}{\partial z} > 0$ (function $f_{\text{lse}}(x)$, blue curve in the left panel Fig. 2 is positive for $x > 0$), which is unphysical. The boundary layer approximation in Eq. 4 (function $f_{\text{bl}}(x)$, red curve in the left panel) is, however, zero for $x > 0$, which is physically correct.

It should be observed that the log-sum-exp approximation gives a better approximation with $\mu \gg 5$, but Open-Modelica fails to linearize the model when $\mu \gtrsim 5$.

2.3 Transport delay in heating loop

The advection term related to flow of water in the loop of the heater and the floor, has been approximated by a Padé approximation as in (Bhattarai et al., 2020):

$$\frac{T(s; x = L)}{T(s; x = 0)} = \frac{1}{\exp(\tau_N s)} \prod_{i=1}^{N-1} \frac{\exp\left(-\frac{\tau_i}{2}s\right)}{\exp\left(\frac{\tau_i}{2}s\right)}$$

$$\approx \frac{1}{1 + \tau_N s} \prod_{i=1}^{N-1} \frac{1 - \frac{\tau_i}{2}s}{1 + \frac{\tau_i}{2}s}.$$

The advantage of including the lag term $\frac{1}{1+\tau_N s}$ is that this removes the need to differentiate the input signal $T(t; x = 0)$.

3 Simulation with buoyancy approximations

The heated tank model from (Johansen et al., 2019) has been corrected as discussed (Bhattarai et al., 2020). Here, we study the accuracy of the log-sum-exp approximation
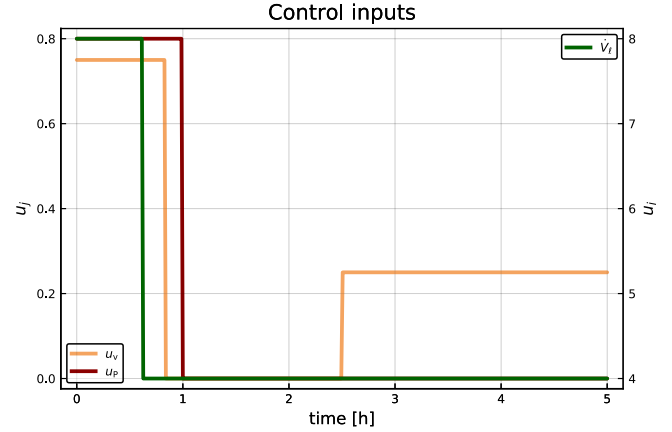
Figure 3. Control inputs u_P (power fraction to heated tank), u_v (water flow valve opening), and $\dot{V}_\ell$ (volumetric flow rate in heating loop).
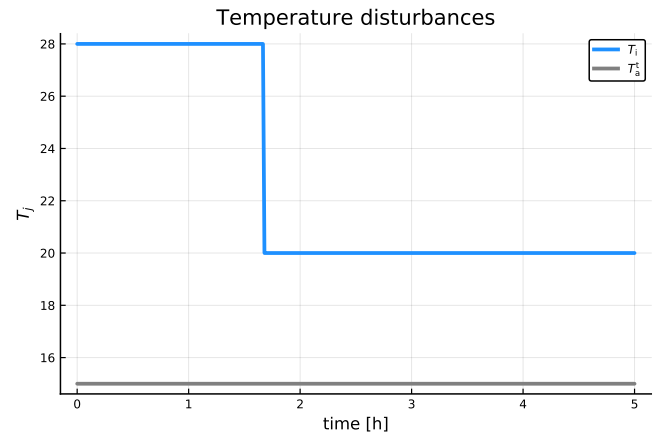
Figure 4. Disturbance temperature inputs T_i (influent to heated tank) and T_a^t (ambient to tank).

and the boundary layer approximation introduced in Section 2.2, when applied to the heated tank. Next, we study the effect of the approximations on a combined heater and floor heating loop, as discussed in (Bhattarai et al., 2020). In all cases, stratification models with 20 discretization layers are used in the heated tank models.

3.1 Heated tank

The following input signals are used for the heated tank: Fig. 3 depicts the power input u_P, the valve input u_v, and the volumetric loop flow rate $\dot{V}_\ell$.

Figure 4 shows the input temperature T_i to the heated tank, and the ambient temperature surrounding the heater, T_a^t.

The temperature distribution in the heated tank using the corrected expression for buoyancy conduction compared to (Johansen et al., 2019), with approximations, is shown in Fig. 5 when using OpenModelica with solver DASSL. In Fig. 5, observe in particular that the log-sum-exp approximation gives a different solution for some internal heated tank locations, see ellipses I and II in the figure.
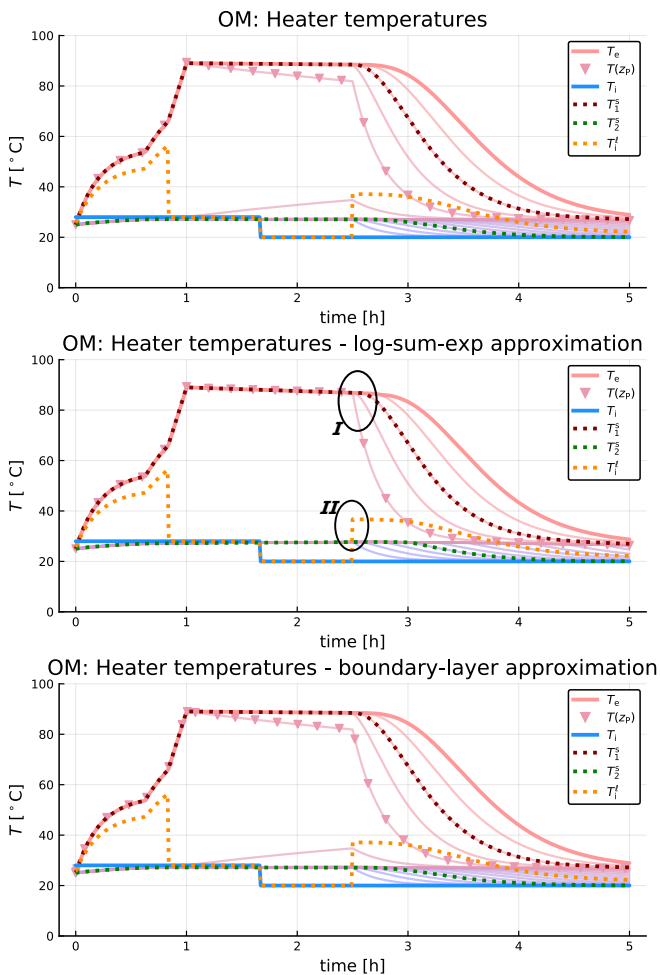
Figure 5. Temperature distribution (T_e: effluent temperture from heated tank; $T(z_P)$: temperature at heating element; T_i: influent temperature to heated tank; T_j^s temperature at heated tank sensors; T_i^ℓ inlet temperature to floor heating loop) in tank with corrected buoyancy conduction + approximations. Model implemented in Modelica, solved via OMJulia using OpenModelica with default solver `DASSL`.
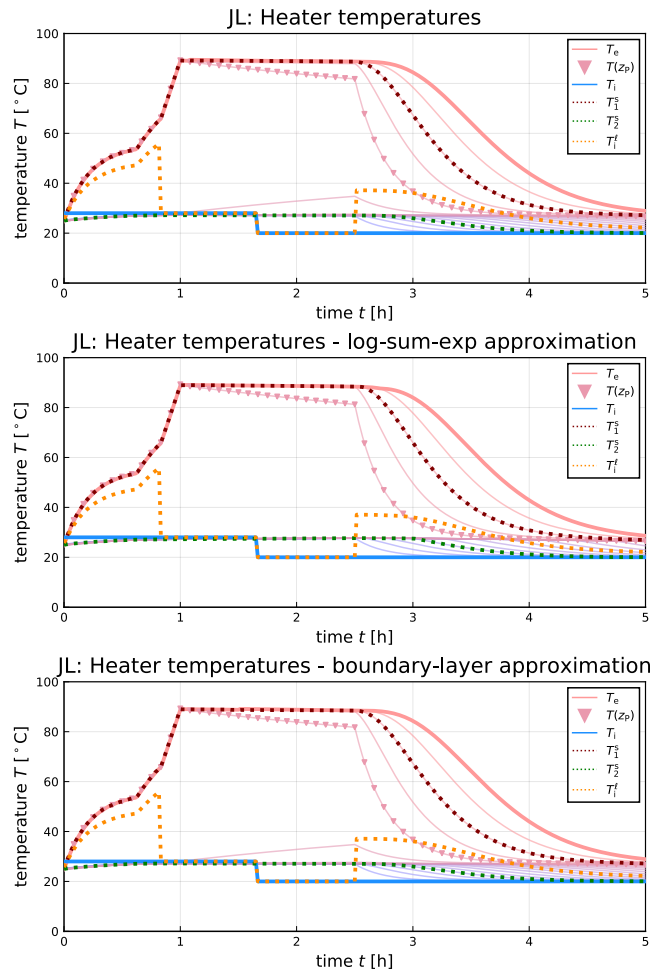
Figure 6. Temperature distribution in tank with corrected buoyancy conduction + approximations. Model implemented in Julia, solved via package `DifferentialEquations` with solver `CVODE_Adams` from the `Sundials` package.

The similar simulations in Julia are depicted in Fig. 6. It turns out that most Julia `DifferentialEquations` solvers struggled with solving the log-sum-exp approximation, reporting that the system is unstable. The `CVODE` solvers of the `Sundials` package handles the problem, though, and the `CVODE_Adams` solver appears to give the solution closest to the original formulation with the max function. Observe that the log-sum-exp approximation in Julia avoids the problem associated with ellipse I of Fig. 5, but retains the problem of ellipse II of the OpenModelica solution. However, for Julia, the expression `log(1+x)` can be replaced by an improved function `log1p(x)` when `x` is small. When doing so, the problem of ellipse II is also removed, and the log-sum-exp approximation can be made virtually indistinguishable from the solution of the max function in Julia. Still, both the original max formulation and boundary layer approximation allows for the use of standard solvers, while the log-exp-sum approximation requires `Sundials` solvers to find the solution.

Both when using the OpenModelica solver and the Julia solver, the boundary layer approximation works without
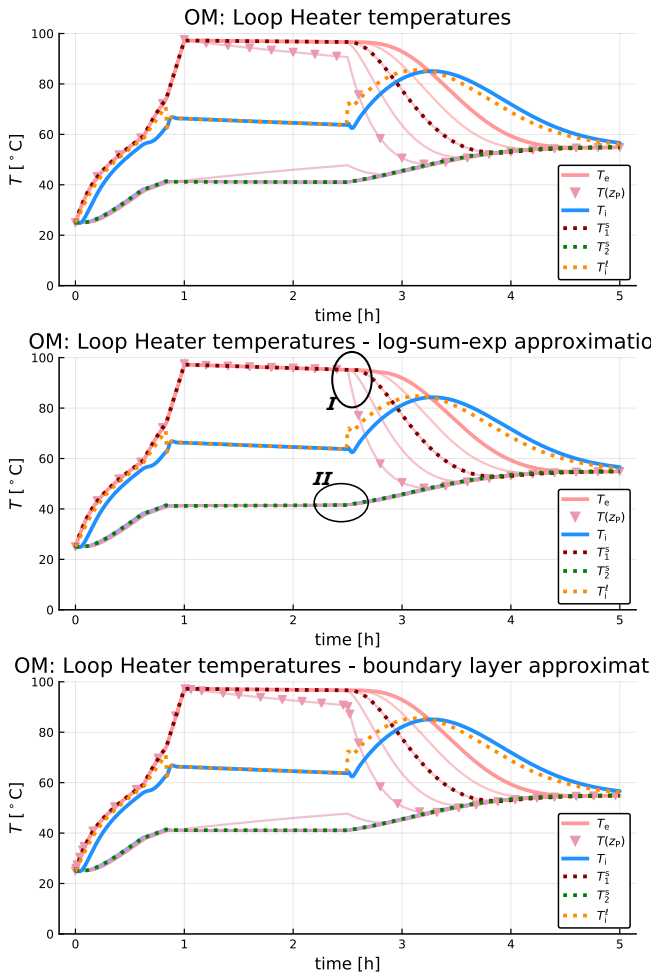
Figure 7. Temperature distribution in tank with floor loop and with corrected buoyancy conduction + approximations. Model implemented in Modelica, solved via OMJulia using OpenModelica with default solver `DASSL`.

problems, and gives a solution very similar to the original solution. In summary, the results indicate that the log-sum-exp approximation is numerically challenging, even when using such a "poor" value of μ as $\mu = 5$, see Fig. 2.

3.2 Heated tank + floor heating loop

Next, we consider the combined heated tank and floor heating loop. The inputs are as in Figs. 3–4, except that the input temperature T_i (Fig. 4) to the heated tank now is a state, and is computed from the model.

The temperature distribution in the heated tank with floor loop using the corrected expression for buoyancy conduction compared to (Johansen et al., 2019) as well as approximations, are shown in Fig. 7.

Again, Fig. 7 indicates some numeric problem with the log-sum-exp approximation, see ellipses I and II.

It is also of interest to see whether there are differences in the floor-room temperatures. The temperatures in the floor layers and room of the tank with floor loop (original buoyancy expression and approximations) are shown in Fig. 8.
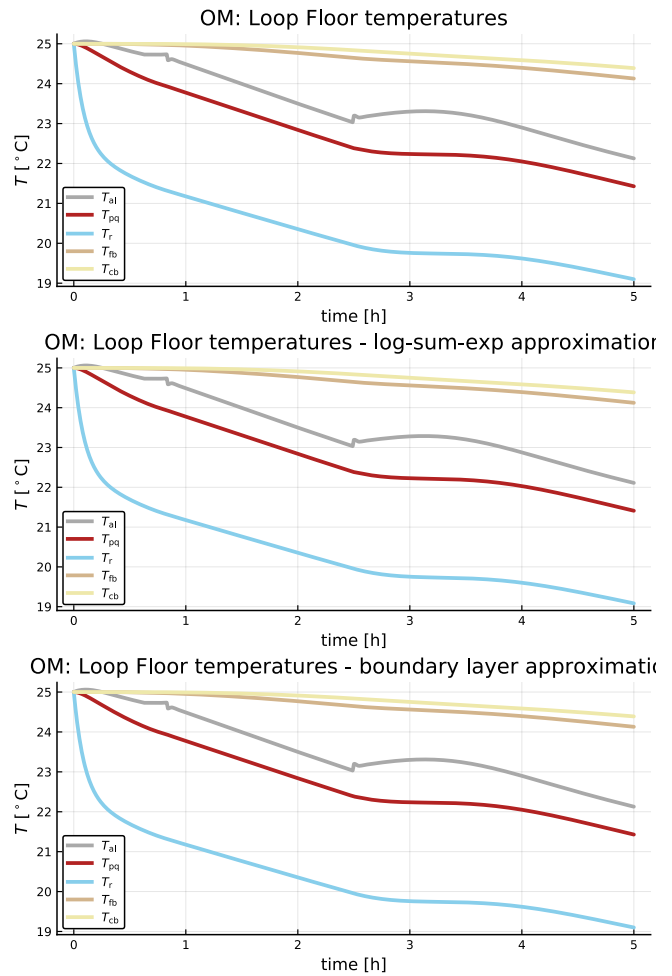
Figure 8. Temperature distribution in floor layers and room of heated tank with floor loop (T_j: temperature in compartment j — al: aluminum plate, pq: parquet, r: room, fb: fiber board, cb: chip board). Model implemented in Modelica, solved via OMJulia using OpenModelica with default solver `DASSL`.
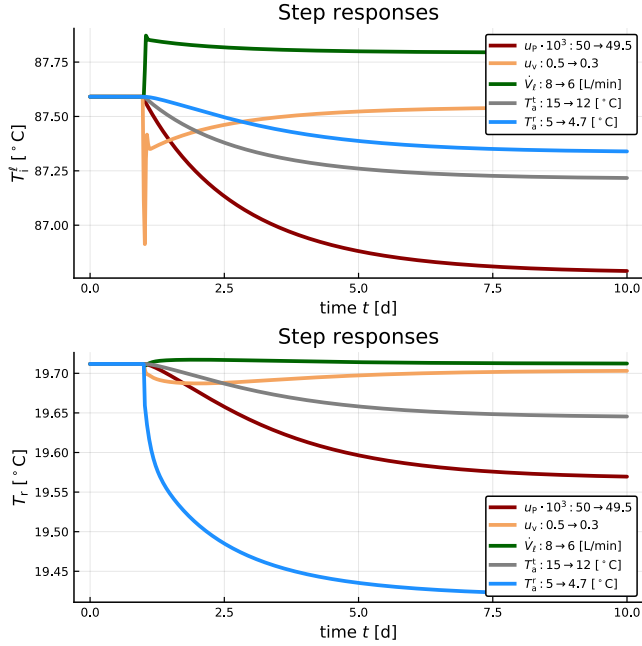
Figure 9. Step responses in T_i^ℓ (upper panel) and T_r (lower panel) to changes in inputs (u_P: power fraction in heater, u_v: heater bypass fraction, $\dot{V}_\ell$: loop volumetric flow rate, T_a^t: ambient temperature for heated tank, T_a^r: ambient temperature for room, i.e., outdoor temperature), starting at steady state.

As seen from Fig. 8, the temperatures in the floor layers and the room are practically the same, independently of the buoyancy approximation.

4 Model analysis

4.1 Step response

The response in an output to a step change in an input, starting from steady state, gives a good idea of time constants in the system. Figure 9 shows the responses in temperatures T_i^ℓ and T_r to step changes in inputs.

As seen Fig. 9, a step change in the heater by-pass fraction (u_v) gives an inverse response in both the loop inlet temperature (T_i^ℓ) and the room temperature (T_r). Inverse response is intimately related to right half plate (RHP) zeros, and indicate a limitation in how fast the system can be controlled using this control signal. A similar inverse response may be seen in the response to $\dot{V}_\ell$, while the response to u_P does not have inverse response. Based on these observations, u_P is probably the best choice of control signal.

Step changes in disturbances T_a^t and T_a^r are related to similar limitations in observers/state estimators, but in that case, the assessment is more complicated.

4.2 Parameter sensitivity

Sensitivity in output y to a parameter θ is defined as $\frac{\partial y}{\partial \theta}$, and is informative as to what degree an infinitesimal change in θ will change the output y. If $\frac{\partial y}{\partial \theta} \equiv 0$, this implies that changing parameter θ will not change the model
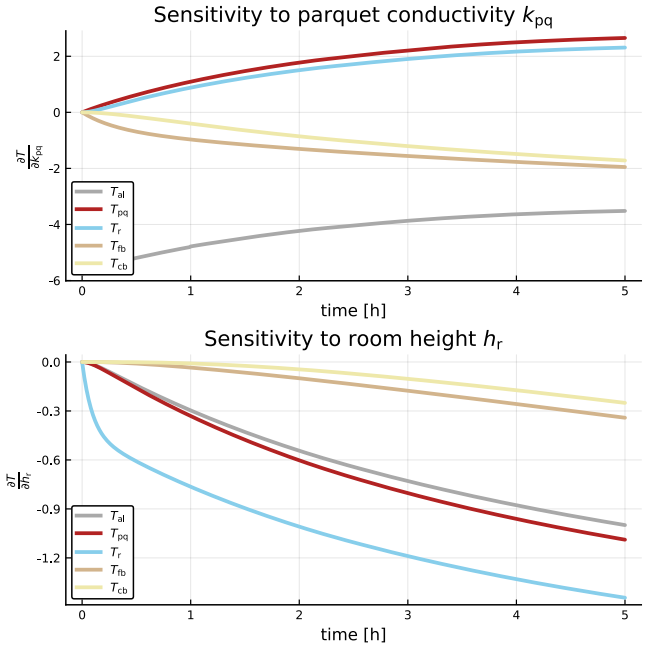
Figure 10. Sensitivities in floor-room temperatures to infinitesimal changes in parquet conductivity k_pq and height of room, h_r.

output y at all, hence parameter θ can not be used for improving the model fit. On the other hand, if $\frac{\partial y}{\partial \theta} \neq 0$, this indicates that changing θ will change the output y and allow for improved model fit. The transient change of $\frac{\partial y}{\partial \theta}$ is also informative wrt. what parts of the model parameter θ can be used to improve, e.g., steady state value, time constants, etc.

Both OMJulia-OpenModelica and the DifferentialEquation package for Julia support sensitivity computations. Here, we illustrate the results of sensitivity computations in OMJulia-OpenModelica for $y = \left(T_\mathrm{al}, T_\mathrm{pq}, T_\mathrm{r}, T_\mathrm{fb}, T_\mathrm{cb}\right)$ and $\theta = \left(k_\mathrm{pq}, h_\mathrm{r}\right)$. In the sensitivity computations, we need to specify inputs. We start with inputs $u = \left(\begin{array}{ccccc} T_\mathrm{a}^\mathrm{r} & T_\mathrm{a}^\mathrm{t} & \dot{V}_\ell & u_\mathrm{P} & u_\mathrm{v} \end{array}\right)$ given by nominal values

$$u^* = \left(\begin{array}{cccc} 5\,^\circ\mathrm{C} & 15\,^\circ\mathrm{C} & 5\,\mathrm{L/min} & 1.25\cdot10^{-2} & 1 \end{array}\right)$$

at $t = 0$ and corresponding steady states, and inject a step change in u_v from $u_\mathrm{v} = u_\mathrm{v}^* = 1$ to $u_\mathrm{v} = 0.8$ at time $t = 1\,\mathrm{h}$. Sensitivities of model `fhm` are computed numerically via OMJulia command

```julia
julia> sn,sv = sensitivity(fhm,["kpq","hr"],[
    "Tal","Tpq","Tr","Tfb","Tcb"])
```

The results are displayed in Fig. 10.

To properly assess parameter identifiability, one needs to consider the rank of the sensitivity matrix, see, e.g., (Sarmiento Ferrero et al., 2006).

4.3 Poles and zeros

OMJulia, the Julia API to OpenModelica, allows for automatic differentiation for linearization of models. The algorithm does not handle the original buoyancy conduction
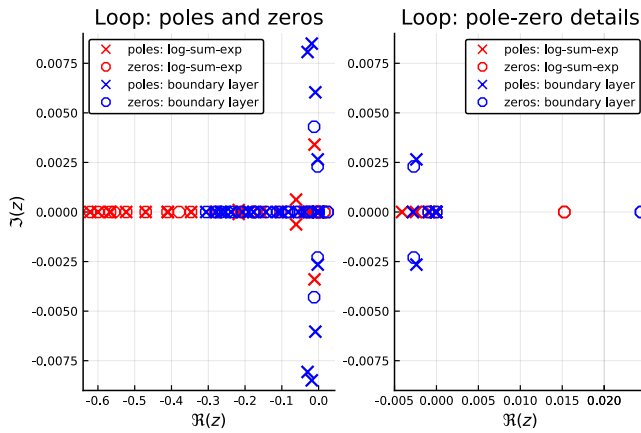
Figure 11. Poles marked with × (red: log-sum-exp approximation, blue: boundary layer approximation), and transmission zeros marked with ○.

Figure 12. Bode plot of transfer function from outdoor temperature T_a^r to room temperature T_r.

Figure 13. Bode plot of transfer function from heater split range valve signal u_v to parquet temperature T_{pq}.

expression due to the non-differentiability of k_b at $\frac{\partial T}{\partial z} = 0$, but works with the two approximate conductivity expressions. In this section, we consider inputs u and outputs y defined as:

$$u = \begin{pmatrix} T_a^r & T_a^t & \dot{V}_\ell & u_P & u_v \end{pmatrix}$$
$$y = \begin{pmatrix} T_i & T_i^\ell & T_{pq} & T_r \end{pmatrix}$$

with nominal input values

$$u^* = \begin{pmatrix} 5°C & 15°C & 5L/min & 1.25 \cdot 10^{-2} & 1 \end{pmatrix}.$$

Linearization of model `fhlse` is carried out and piped (`|>`) into a Julia state space object `sys_lse` via OMJulia+Julia ControlSystems command

```
1  julia> sys_lse = linearize(fhlse) |> x -> ss(
       x...)
```

When starting the model at steady state for the nominal inputs u^*, the combined heater + floor heating system gives a linear approximation with poles and transmission zeros as in Fig. 11.

As seen, there is some difference in both poles and transmission zeros of the two approximations. The right panel of Fig. 11 clearly indicates that both buoyancy approximations exhibit right-half-plane *zeros* in the complex plane, which indicates limitations on how fast the system can be controlled. These right-half-plane zeros are due to advection in the floor heating loop, and the Padé approximation that has been applied. It is difficult to see from Fig. 11 whether all poles of the system lie in the left-half-plane of the complex plane, but the simulations show that the system in fact is stable.

The precise location of *transmission* zeros gives additional information related to that of a possible inverse response found through step responses, see, e.g., (Lie, 1995). Observe that the two buoyancy conductivity approximations give different results wrt. poles and zeros. It is believed that the boundary layer approach is more accurate than that of the log-sum-exp approach.

4.4 Bode plots

It is of interest to also compare the Bode plots from u to y of the linear approximations of the two buoyancy conduction approximations. In this case, we consider the following inputs u and outputs y:

$$u = \begin{pmatrix} T_a^r & \dot{V}_\ell & u_v \end{pmatrix}$$
$$y = \begin{pmatrix} T_i^\ell & T_{pq} & T_r \end{pmatrix}.$$

The reason why we do not consider input u_P is that it turns out that the outputs are insensitive to u_P. The following selected Bode plots are representative of the similarity of the two approximations. Transfer function $\frac{T_r}{T_a^r}(s)$, Fig. 12.

Observe that the two approximations are overlapping. Transfer function $\frac{T_{pq}}{u_v}(s)$, Fig. 13.

Bode plots are standard tools for tuning SISO PID controllers. The two plots in Figs. 12–13 are representative of the variation in similarity between Bode plots from the boyancy conductivity approximations; for a real control design, the Bode plot from u_P to T_r would be more relevant, but controller design is not the main focus of this paper; rather the focus is on tools to analyze models.

5 Conclusions

A model for floor heating of buildings is studied wrt. control relevant properties. The main emphasis is not on the analysis itself, but rather on the use of modern simulation tools to carry out such analysis. For a more realistic analysis, a better balanced system should be considered (size of heater vs. heat loss in room) with improved description of the living space (e.g., more rooms, thermal mass of furniture, ventilation, etc.).

The results indicate the importance of a suitable model description which is differentiable (stratification): this is important both for linearized analysis and controller design, but a proper choice of approximation will also allow for more flexibility in choice of differential equation solvers. The log-sum-exp approximation is locally unphysical, which probably also is the reason why this approximation requires very good solvers.

The results also indicate the ease at which control analysis can be carried out. Step responses are important for assessing overall open loop time constants, as well as indicating the presence of inverse response/"unstable" system zeros. Sensitivity analysis is important for assessing parameter identifiability. The location of open loop poles is related to open loop time constants, while system zeros give crucial information about how fast a closed loops system can be made. Finally, a Bode plot is often used for simple design of linear controllers.

In reality, temperature control in buildings is often done using thermostats and other nonlinear controllers. Still, it is of interest to consider how a more advanced controller — even a linear one — can be utilized for improved use of energy in building temperature control. The methods discussed in this paper constitute a first step in such a control design.

References

Karl Johan Åstrøm and Richard M. Murray. *Feedback Systems. An Introduction for Scientists and Engineers*. Princeton University Press, Princeton, NJ, 2008. ISBN 978-0-691-13576-2.

Ashish Bhattarai, Zahir Barahmand, Sina Orangi, and Bernt Lie. Extended model for control of thermal energy in buildings. In *Submitted to SIMS 2020*, 2020.

Peter Fritzson. *Principles of Object-Oriented Modeling and Simulation with Modelica 3.3: A Cyber-Physical Approach*. Wiley-IEEE Press, Piscataway, NJ, second edition, 2015. ISBN 978-1-118-85912-4.

N.S. Jayakumar, A. Agrawal, M.A. Hashim, and J.N. Sahu. Experimental and theoretical investigation of parametric sensitivity anddynamics of a continuous stirred tank reactor for acid catalyzedhydrolysis of acetic anhydride. *Computers and Chemical Engineering*, 35:1295–1303, 2011.

Casper Amandus Johansen, Bernt Lie, and Nils-Olav Skeie. Models for control of thermal energy in buildings. In Erik Dahlquist, Esko Juuso, Bernt Lie, and Lars Eriksson, editors, *Proceedings of the 60th Conference on Simulation and Modelling (SIMS 59)*, number 170 in Linköping Electronic Conference Proceedings. Linköping University Electronic Press, 2019. doi:https://doi.org/10.3384/ecp20170.

Huibert Kwakernaak and Raphael Sivan. *Linear Optimal Control Systems*. Wiley-Interscience, New York, 1972.

Jesus Lago, Fjo De Ridder, Wiet Mazairac, and Bart De Schutter. A 1-dimensional continuous and smooth model for thermally stratified storage tanks including mixing and buoyancy. *Applied Energy*, 248:640–655, 2019.

Bernt Lie. *Methods of Model Based Process Control*, chapter Attainable Performance in LQG Control, pages 263–295. NATO ASI Series, Series E: Applied Sciences — Vol. 293. Kluwer Academic Publishers, Dordrecht, 1995. Proceedings of the NATO Advanced Study Institute on Methods of Model Based Process Control, Antalya, Turkey, August 7–17, 1994.

Bernt Lie, Carlos Pfeiffer, Nils-Olav Skeie, and Hans-Georg Beyer. Models for solar heating of buildings. In Alireza Rezania Kolai, Kim Sørensen, and Mads Pagh Nielsen, editors, *Proceedings, 55th International Conference of Scandinavian Simulation Society*, pages 28–38, Aalborg University, Denmark, October 2014. Scandinavian Simulation Society, Linköping University Electronic Press. doi:www.ep.liu.se/ecp/108/ecp14108.pdf. October 21–22 2014.

Bernt Lie, Arunkumar Palanisamy, Alachew Mengist, Lena Buffoni, Martin Sjölund, Adeel Asghar, Adrian Pop, and Peter Fritzson. OMJulia: An OpenModelica API for Julia-Modelica Interaction. In *Proceedings of the 13th International Modelica Conference*, pages 699–708, February 2019. doi:10.3384/ecp19157. Regensburg, Germany, March 4–6, 2019.

Christopher Rackauckas and Qing Nie. DifferentialEquations.jl — A Performant and Feature-Rich Ecosystem for Solving Differential Equations in Julia. *Journal of Open Research Software*, 5(15), 2017. doi:10.5334/jors.151.

Cristina Sarmiento Ferrero, Qian Chai, Marta Dueñas Díez, Sverre H. Amrani, and Bernt Lie. Systematic analysis of parameter identifiability for improved fitting of a biological wastewater model to experimental data. *Modeling, Identification and Control*, 27(4):219–238, 2006. doi:10.4173/mic.2006.4.2.

Arvind Varma, M. Morbidelli, and H. Whu. *Parametric Sensitivity in Chemical Systems*. Cambridge University Press, 1999. ISBN 978-0521621717.

R. Viskanta, M. Behnia, and A. Karalis. Interferometric observations of the temperature structure in water cooled or heated from above. *Advances in Water Resources*, 1(2):57–69, 1977.

Evangelos Vrettos. *Control of Residential and Commercial Loads for Power System Ancillary Services*. PhD thesis, ETH, Power Systems Laboratory, 2016.

Zhijie Xu, Ruisheng Diao, Shuai Lu, Jianming Lian, and Yu Zhang. Modeling of electric water heaters for demand response: A baseline pde model. *IEEE Transactions on Smart Grid*, 5(5):2203–2210, September 2014. doi:10.1109/TSG.2014.2317149.

Numerical analysis of the performance of green façades

Frédéric Poupart   Oda Martine Sundsdal   Marianne Christiansen   Ole Melhus   Arnab Chaudhuri

Department of Civil Engineering and Energy Technology
OsloMet — Oslo Metropolitan University, Oslo, Norway.
`frederic.poupart@hotmail.com,oda-martine@hotmail.com, mchrist@hotmail.no,`
`ole.melhus@oslomet.no, arnab.chaudhuri@oslomet.no`

Abstract

This study reports a numerical analysis of the performance of green façades in different geographical locations and seasonal conditions. A mathematical model from a previous study is implemented and combined with the modified convective heat transfer coefficients from a recent study of the literature to simulate the transient heat transfer through bare walls and green facades with climbing vegetation. An implicit Finite Difference Method (FDM) based solver is used to perform the numerical simulations. Climate data are taken from relevant weather stations in Oslo and Rome and typical meteorological year (TMY) values are used for this purpose together with variable thermo-physical properties of air. An energy budget analysis reveals that the short-wave radiation term and convective heat transfer term are predominating compared to the other terms involved in the energy balance equation for summer time. The results show that the green walls are most effective in summer seasons with high levels of solar radiation, as most of the cooling effect is credited to the vegetation blocking the solar radiation. In cooler seasons, the benefit is less prominent. Furthermore, an analysis of the effects of the different models of convective heat transfer coefficients is presented.
Keywords: green buildings, green vertical systems, green façades, passive design, sustainability

1 Introduction

Recent studies show a drastic increase of global energy use over the last four decades. Based on the United Nations report and as mentioned by several authors, it is estimated that the population in cities will increase up-to 68% by 2050 (UN, 2018; Vo et al., 2019; Seyam, 2019; Besir and Cuce, 2018) with urban dwellers reaching ≈ 6.7 billion. In this regard, the building sector accounts for a large portion of the total energy usage, as well as 36% of the total greenhouse gas emissions is attributed to the building sectors (Vo et al., 2019; Seyam, 2019). For this, the United Nations Environment Program (UNEP) proposes several measures like increasing the energy efficiency of buildings, appliances, reduction of emission, approach towards energy usage and promoting technologies with renewable energies rather than fossil fuels.

Integrating plants into the building envelope addresses both aesthetic and energy concerns. Utilizing different types of greenery systems, microclimatic conditions of buildings can be more cost-effective and eco-friendlier (Seyam, 2019; Koc et al., 2018; Besir and Cuce, 2018). Green buildings with green infrastructures like green horizontal systems and green vertical systems are sustainable, efficient and passive design solutions for effects associated with rapid urbanization, addressing issues with a high carbon footprint, high greenhouse gas emission, urban heat island effects etc. (Susca et al., 2011; Safikhani et al., 2014; Manso and Castro-Gomes, 2015; Charoenkit and Yiemwattana, 2016; Koc et al., 2018). Green roofs (GR) and vertical greenery systems like green façades(GF)/living walls differ fundamentally due to the structure of the assembly and differences in the involved heat transfer processes (Susorova et al., 2013; Pérez et al., 2014; Bustami et al., 2018). A detailed classification of green walls based on the construction characteristics is presented by (Manso and Castro-Gomes, 2015). Scientific studies established the energy saving benefits of green infrastructures in building sectors (Feng and Hewage, 2014; Raji et al., 2015). It is also a proven fact that the presence of green systems has an important psychological impact on urban dwellers. Thermal insulation, shading effects, evapotranspiration and wind effects are the key features towards the energy saving impacts of greenery systems. Several parameters like substrate thickness, soil moisture, solar radiation inception, density of foliage and the coverage, ventilation blind effects, thickness of the foliage, air layer, orientation, wind barrier/blockage etc. are to be accounted for in mathematical modeling of such systems. Plant characteristics, weather conditions, climate zones, wall assembly types, façade orientations etc. are to be considered in modeling and analysis of thermal benefits.

In this regard, Alexandri and Jones developed a dynamic one dimensional model for green roof accounting the fundamental heat and mass transfer mechanisms (Alexandri and Jones, 2007) involved in the energy dynamics. On the other hand, a quasi steady state approach was adopted by (Tabares-Velasco and Srebric, 2012) in their green roof model. Further developments on the green roof model are reported in (Heidarinejad and Esmaili, 2015; He et al., 2017). For vertical greenery systems, a mathematical model of building exterior wall with climbing vegetation is presented in (Susorova et al., 2013). Sev-

eral other authors considered the heat balance of the foliage and heat transfer through the substrate layer of the vertical greenery system and integrated models in building simulation programs like TRNSYS and EnergyPlus (Djedjig et al., 2015, 2017; Dahanayake and Chow, 2017). A multilayered model approach for vegetated vertical systems is presented in (Convertino et al., 2019a). Nevertheless, studies on the modeling aspects of green roof/green façades are still an active field of research and development.

In this work, we have revisited the model proposed in (Susorova et al., 2013). We aim to implement this green façade model combining the effects of the modified convective heat transfer coefficients presented in the recent work of (Convertino et al., 2019b). The main objective of this work is to investigate the applicability of the aforementioned model and to analyze the performance of the green wall envelope of buildings, for different climate data revealing the possible benefit of employing green façades in buildings. The paper is organized as follows. In section 2, we present the governing equations and the numerical methodology of the study together with the problem setup. The detailed analysis of the results is given in section 3. Finally, the conclusions are drawn in section 4.

2 Method

2.1 Governing equation

The green façade model of Susorova et al. accounts the various heat transfer mechanisms for a vegetated wall (Susorova et al., 2013), and the energy balance is defined as:

$$SR_{vw} + LR_{vw} + XR + C_{vw} = Q_{vw} + S_{vw} \qquad (1)$$

Where SR_{vw} is the incident shortwave radiation, LR_{vw} is the net long-wave radiation, XR is the radiative exchange between the leaves and wall surface, C_{vw} is the convective heat flux, Q_{vw} is the conduction heat flux and the heat storage in the façade material is $S_{vw} = \rho C_p L (dT_{vw}/dt)$, with density of the wall material ρ, specific heat of the wall C_p, wall thickness L, t is the time and T_{vw} is the wall temperature. A schematic of the energy balance and a very brief description of the terms of the Equation 1 are given in Appendix A for the sake of completeness and to facilitate the understanding of the overview of the mathematical model. Details of the model can be found in (Susorova et al., 2013). Note that essentially, the wall temperature $T_{vw}(t)$ is linked with the air temperature $T_{air}(t)$, the sky temperature $T_{sky}(t)$, the ground temperature $T_{gr}(t)$, the indoor air temperature $T_{in}(t)$ and the leaf temperature $T_{leaf}(t)$. The energy balance equation of bare wall is similar to equation 1, without the XR term and the unknown bare wall temperature denoted as T_{bw}. The energy balance equation is therefore expressed as:

$$SR_{bw} + LR_{bw} + C_{bw} = Q_{bw} + S_{bw} \qquad (2)$$

The governing ordinary differential equations (ODEs) can be expressed in the following form:

$$\frac{dT_{vw}}{dt} = a_{vw}(t) + b_{vw}(t)T_{vw} + c_{vw}(t)T_{vw}^4 \qquad (3)$$

$$\frac{dT_{bw}}{dt} = a_{bw}(t) + b_{bw}(t)T_{bw} + c_{bw}(t)T_{bw}^4 \qquad (4)$$

Where, the time varying coefficients (a_{vw}, b_{vw}, c_{vw}, a_{bw}, b_{bw} and c_{bw}) are functions of the other associated temperatures and the relevant thermo-physical properties.

2.2 Numerical procedure

The implicit form of the discretized equations for the ODEs can be given by,

$$\frac{T_{vw}^{n+1} - T_{vw}^n}{\Delta t} = a_{vw}^{n+1} + b_{vw}^{n+1}T_{vw}^{n+1} + c_{vw}^{n+1}(T_{vw}^{n+1})^4 \qquad (5)$$

$$\frac{T_{bw}^{n+1} - T_{bw}^n}{\Delta t} = a_{bw}^{n+1} + b_{bw}^{n+1}T_{bw}^{n+1} + c_{bw}^{n+1}(T_{bw}^{n+1})^4 \qquad (6)$$

Where, Δt is the time step and the superscripts n and $n+1$ denotes the time levels of the discretization.

The implicit FDM based solver is developed using Matlab programming language to compute the transient temperature of the wall with a time step of one minute.

2.3 Problem setup

Simulations are performed in which the temperature of the façade is computed over a 24-hour period. To perform these computations, hourly weather data is gathered through the EU's photovoltaic geographical information system and a local weather station in the relevant cities (Norsk klimaservicesenter, 2020; The European Commission's science and knowledge service, 2019; YAGA AS, 2020). A suitable interpolation procedure is adopted to produce weather data for a time resolution in minutes. The green façade is assumed to be consisted of a brick wall covered with Boston Ivy and the corresponding relevant material properties are used. Among the necessary parameters, ρ, C_p, L and T_{in} are assumed to be $672 kg/m^3$, $468 J/kgK$, $0.2m$ and $295K$ respectively. The choice of T_{in} is based on the recommended operating temperature mentioned in Norwegian standard (TEK17, 2017). All other parameters used in the simulations are summarized in Appendix A. Two seasons (summer and winter) are considered for the performance behavior of the mathematical model in the geographical location of the cities Rome and Oslo. Note that an initial temperature of the wall is required to achieve the time marching solutions. To avoid this initial data affecting the results, simulations are performed over a 48 hour period, of which the last 24 hours are extracted as the result. In the next section, we present the results of the simulations for different scenarios.

3 Results and Discussions

In this section, we first illustrate the relative importance of the individual terms of the energy balance equation. This

is followed by the comparison of the performance of the green façades in different scenarios. Finally, the analysis of the impact of the convective heat transfer coefficient with various empirical models is presented at the end.

3.1 Energy budget

The budget of the terms of the energy equation for bare walls and green walls are calculated after finding the wall temperatures. For this analysis, a typical summer day situation in Oslo region is considered. The contributions of each term in the energy equation are shown in Figure 1 and Figure 2 for a bare wall and green wall, respectively. It is clear from these figures that the contributions from the convection term and the solar short-wave radiation term are dominant from 09:00 hrs to 19:00 hrs for both the walls. For example, at 14:00 hrs the contribution of the short-wave radiation term SR_{vw}, and convection term C_{vw} for the green wall are $\approx 49\%$ and $\approx 40\%$ respectively. On the other hand, for bare wall the contribution of SR_{bw} and C_{bw} are found to be $\approx 50\%$ and $\approx 37\%$. It can be noticed that the values of all budget terms are significantly reduced in the green wall essentially due to the blockage of the radiation and associated lower temperature differences. The contribution of the radiative exchange term XR, for the green wall is found to be relatively less compared to the other terms.
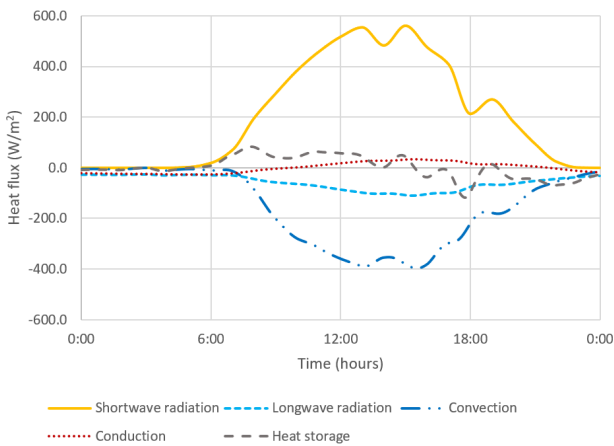
Figure 1. Energy budget bare wall.

3.2 Performance of green façades in different climates and seasons

Here we present the results of the performance of the bare wall and the green wall in two different locations in Europe namely Rome and Oslo using the weather data for the month of July and February 2019.

Figure 3 shows the comparison between the bare wall and green wall temperature in summer situation for both cities. Evidently, green wall reduces the wall temperature and thereby is capable of reducing the cooling demand in peak hours in daytime (6:00 hrs to 18:00 hrs for Rome and 7:00 hrs to 22 hrs in Oslo). This clearly depicts the
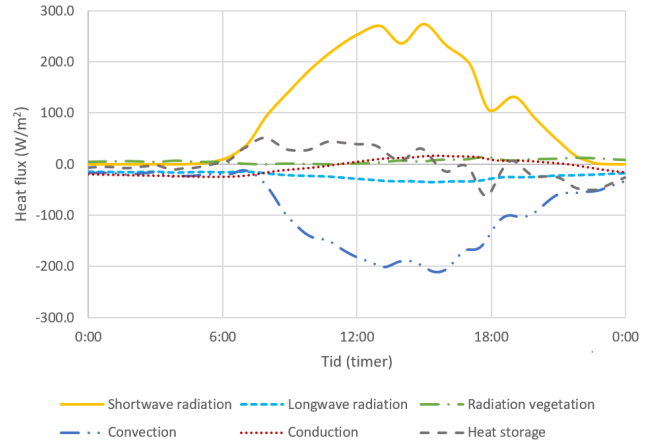
Figure 2. Energy budget green wall.
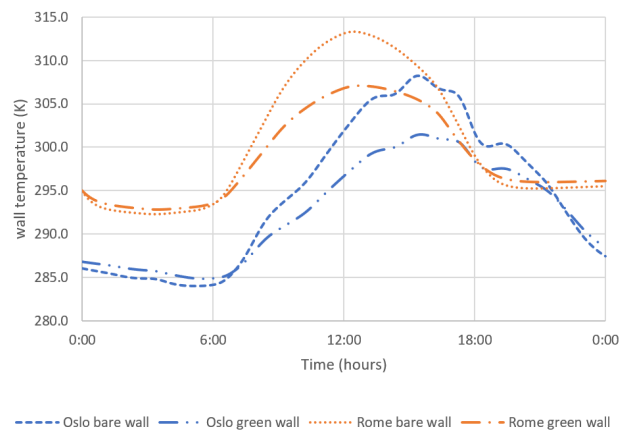
Figure 3. Wall temperature of bare- and green walls over a summer day in Rome and Oslo.

effectiveness of the green vegetation countering with the dominant incident radiation component.

It can be realized that a temperature drop of about $6-7K$ (see Figure 4) is achievable with the application of green vegetation in the summer daytime. Also, the cooling effect lasts relatively longer in Oslo (≈ 3 hrs more) than Rome. This is directly linked with relatively longer daytime in Oslo than Rome in the month of July. Note that, an estimate of the reduction of cooling demand can be made from the conductive heat flux term of the energy balance equation. This cooling demand reduction on an average is found to be $\approx 10W/m^2$ for Rome and $\approx 9W/m^2$ for Oslo. In nighttime, the green wall temperatures are found to be marginally higher than the bare walls in both the cities and the green layer reduces the heat loss due to long-wave radiation.

The results of the simulations in winter scenarios for Rome and Oslo are shown in Figure 5 in terms of wall temperature for both the bare wall and the green wall. In Rome during the daytime, 8:00 hrs to 15:00 hrs a maximum drop of about $2-3K$ (see the temperature difference between the bare wall and the green wall Figure 6) is ob-
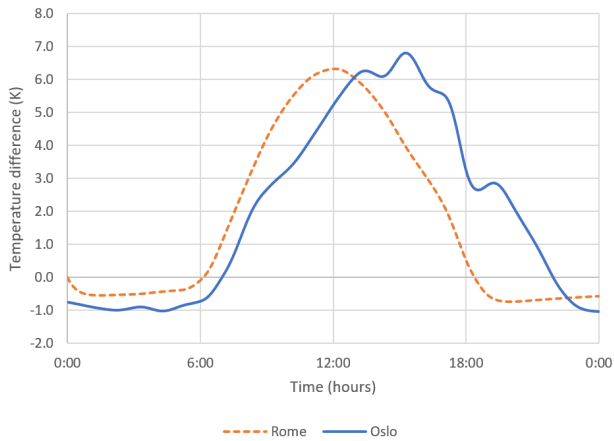
Figure 4. Temperature difference between bare- and green walls over a summer day in Rome and Oslo.

Figure 6. Temperature difference between bare- and green walls over a summer day in Rome and Oslo.
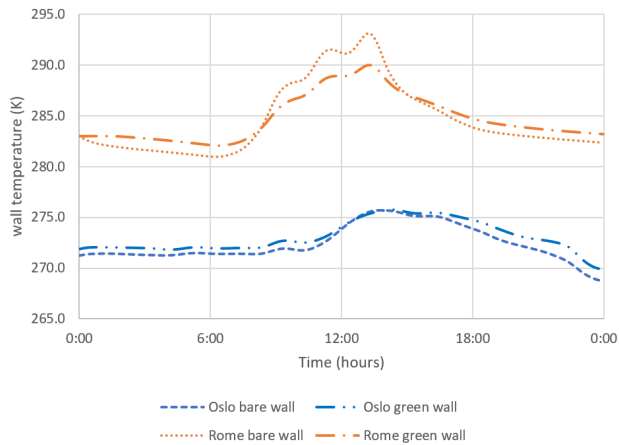
Figure 5. Wall temperature of bare- and green walls in Rome and Oslo, Winter.

served. This leads a marginal increase in heating load during this time. The rest of the day and nighttime the wall temperature remains about $1K$ higher for the green wall. However, in Oslo climate, the winter is much more severe compared to Rome. The green wall temperature remains $\approx 1K$ higher most of the time, except a very short period of the daytime. This provides very marginal benefits on heating demand in places having colder climate.

3.3 Impact of convective heat transfer coefficient

The analysis presented in the previous sections 3.1 and 3.2, are completely based on the mathematical model of Susorova et al. (Susorova et al., 2013). The energy budget analysis in section 3.1 reveals that the convective heat flux term is one of the important contributors in the energy balance equation. The convective heat transfer coefficient for green wall in the energy balance model (Susorova et al., 2013) is approximated with the same correlation as for bare wall case ($h_{vw} \approx h_{bw} = 10.79 + 4.192V$, where V is the air speed at the bare façade). In this work, we intend to
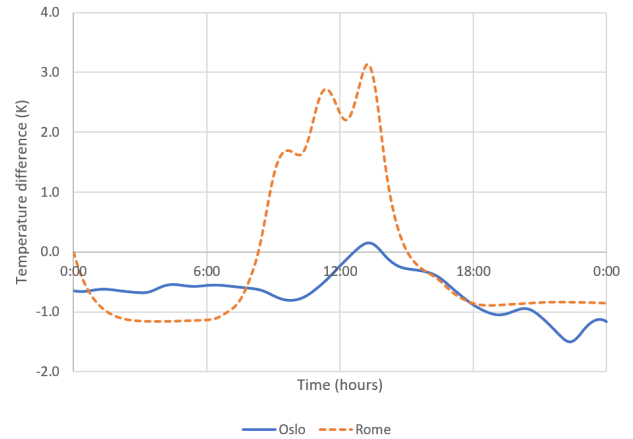
explore the other relevant correlations (Deardorff, 1978; Stanghellini, 1987; Morrison Jr and Barfield, 1981; Ayata et al., 2011; Campbell and Norman, 2012; ASHRAE, 2020) as mentioned in the recent work of Convertino et al. (Convertino et al., 2019b). The expressions of convective heat transfer coefficients are summarized in Appendix B.

A typical summer scenario in the Oslo climate region is considered for this analysis. Figure 7 shows the wall temperature of the green wall for various h_{vw} values as mentioned before. Based on the input weather data, thermophysical properties and plant characteristics we found that the Richardson number $Ri < 10^{-2}$ throughout the day. This implies that the convection type can be considered as forced convection. The application of mixed regime Nusselt number Nu, following (Stanghellini, 1987) is therefore omitted for comparison in this analysis. Note that, increase in h_{vw} leads to decrease in wall temperature comparatively.
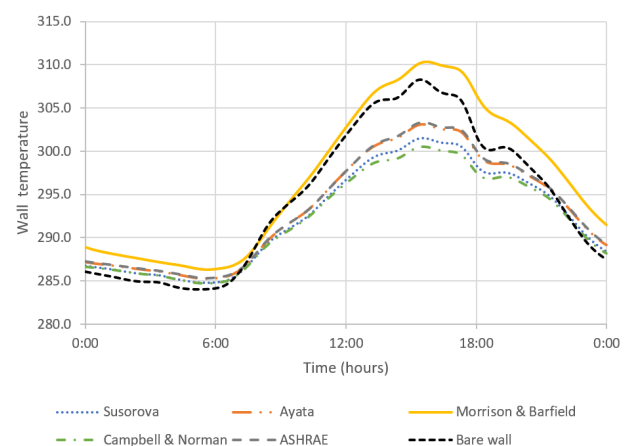
Figure 7. Calculated wall temperature with different approaches of finding the heat transfer coefficient.

It can be seen from Figure 7 that (Campbell and Norman, 2012) approach yields lower wall temperature than that of (Susorova et al., 2013). On the other hand, higher

wall temperatures are obtained by (Ayata et al., 2011; ASHRAE, 2020; Morrison Jr and Barfield, 1981) methods. Among these, (Ayata et al., 2011; ASHRAE, 2020) methods yield almost identical result as they are both based on McAdams' equation . On the other hand, (Morrison Jr and Barfield, 1981) approach gives far lower values for the convection coefficient and therefore much higher wall temperature (even higher than the bare wall) is estimated. It is clear that, there is a wide variation in the prediction of green wall temperature when applying the different methods of calculating the convective heat transfer coefficient compared to Susorova's model. As mentioned in (Susorova et al., 2013), some studies showed a decrease in values of convective heat transfer coefficients for a wide range of vegetated façade types, the present findings through the (Ayata et al., 2011; ASHRAE, 2020) methods are in accordance with this trend. It is known that the building geometry, surroundings and wind directions have a significant effect on the magnitude of the convective heat transfer coefficient (Defraeye et al., 2011; Iousef et al., 2019). Several correlations are based on the results of flat plate experiments in wind tunnels (Jayamaha et al., 1996) and may not fully capture the local variations of geometry and wind directions around the building façades.

4 Conclusions

Modeling aspects of green roof/green façades require to deal with complex interactions associated with plants, soils, building materials, surrounding atmosphere and solar radiation. This fundamentally involves transient coupled heat and mass transfer mechanisms. In this work, we have revisited the mathematical model of building exterior wall with climbing vegetation as presented in (Susorova et al., 2013). An implicit FDM based solver is developed to perform simulations for two different seasons in Oslo and Rome and typical meteorological year (TMY) values are used together with variable thermo-physical properties of air.

An energy budget analysis reveals that the short-wave radiation term and convective heat transfer term are predominating compared to other terms involved in the energy balance equation for summer time. The performance of green wall showed capability of decreasing the cooling demand in summer time quite significantly for both Rome and Oslo. With recent trends in climate change, the cooling need in the Oslo area will most likely increase in the next few years. During the winter, however, the decrease in heating demand is not as significant as the reduced cooling demand noticed in summer time. Nevertheless, especially in relatively cold winter regions like Oslo, the green wall still helps to increase the wall temperature compared to the bare wall. Furthermore, we have analyzed several different methods for the estimation of the convective heat transfer coefficient as the convective term in the energy balance equation plays an important role opposing the solar radiation gain. The prediction of the behavior of the

wall temperatures for green wall with methods in previous literatures are in accordance with the expected trend. Nevertheless, further detailed studies with supporting experimental data, the limitations of this simplified green façade model can be improved. The detailed analysis of the longwave radiation term, radiative exchange term between the leaves and the wall surface will be undertaken in our future work.

Appendix A

In this section, we present a brief description of the mathematical model (see details in (Susorova et al., 2013)) and the input parameters used in the simulations. Figure 8 illustrates the schematic view of the green wall setup and associated energy transfer processes.
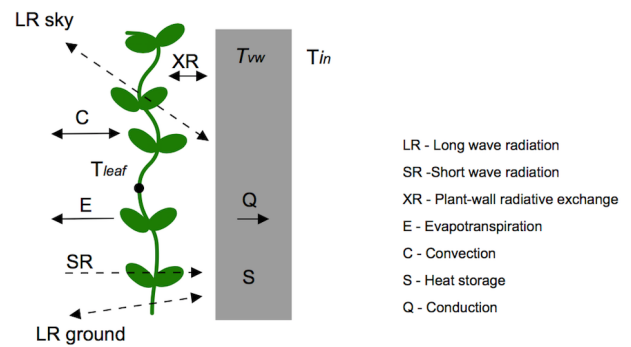
Figure 8. Schematic of the energy balance of the green wall (Susorova et al., 2013; Vo et al., 2019).

The terms of the equation 1 for green wall are given by,

$$
\begin{aligned}
SR_{vw} &= I_t \alpha_w \tau \\
LR_{vw} &= \tau \varepsilon_w \varepsilon_{sky} \sigma F_{sky}(T_{sky}^4 - T_{vw}^4) + \tau \varepsilon_w \varepsilon_{gr} \sigma F_{gr}(T_{gr}^4 - T_{vw}^4) \\
XR &= (1-\tau)\frac{\varepsilon_w \varepsilon_{leaf} \sigma}{\varepsilon_w + \varepsilon_{leaf} - \varepsilon_w \varepsilon_{leaf}}(T_{vw}^4 - T_{leaf}^4) \\
C_{vw} &= h_{vw}(T_{air} - T_{vw}) \\
Q_{vw} &= (T_{vw} - T_{in})/R_{bw} \\
S_{vw} &= \rho C p L (dT_{vw}/dt)
\end{aligned}
$$

On the other hand, the terms of the equation 2 for bare wall are given by,

$$
\begin{aligned}
SR_{bw} &= I_t \alpha_w \\
LR_{bw} &= \varepsilon_w \varepsilon_{sky} \sigma F_{sky}(T_{sky}^4 - T_{bw}^4) + \varepsilon_w \varepsilon_{gr} \sigma F_{gr}(T_{gr}^4 - T_{bw}^4) \\
C_{bw} &= h_{bw}(T_{air} - T_{bw}) \\
Q_{bw} &= (T_{bw} - T_{in})/R_{bw} \\
S_{bw} &= \rho C p L (dT_{bw}/dt)
\end{aligned}
$$

Where I_t is the total solar radiation incident on the wall, α_w is the wall absorptivity, τ is the plant layer transmissivity of radiation, ε_w, ε_{sky}, ε_{gr} and

ε_{leaf} are the emissivities of the wall, sky, ground and plant layer, respectively. The values are assumed as $\alpha_w = 0.7, \varepsilon_w = 0.9, \varepsilon_{sky} = 1, \varepsilon_{gr} = 0.9, \varepsilon_{leaf} = 0.96$. The plant layer transmissivity, $\tau = \exp(-\kappa LAI)$. Here, the radiation attenuation coefficient κ (taken as $0.4W/mK$), and the leaf area index LAI (assumed as 1.8). σ is the Stefan-Boltzmann constant. F_{sky} and F_{gr} are the view factors between the wall, sky and ground, T_{sky} is the temperature of the sky, T_{vw} is the temperature of the vegetated wall, T_{bw} is the temperature of the bare wall, T_{gr} is the temperature of the ground (assumed equal to T_{air}), T_{leaf} is the leaf temperature, h_{vw} and h_{bw} are the convection heat transfer coefficient of the vegetated and bare wall (see Appendix B).

T_{sky} is calculated as a function of the air temperature and dewpoint temperature as:

$$T_{sky} = T_{air} \left(0.8 + \frac{T_{dew} - 273}{250} \right)^{0.25}$$

The view factors are calculated as a function of the tilt angle θ

$$F_{gr} = 0.5(1 - \cos\theta)$$
$$F_{sky} = 0.5(1 + \cos\theta)$$

For the vertical greenery systems the tilt angle is equal to $90°$, resulting in both view factors equal to 0.5.

The equation for calculating T_{leaf} can be found in (Susorova et al., 2013). The necessary parameters are the thermodynamic phsycrometer constant γ ($0.000666°C^{-1}$), radiative conductance through air g_r (obtained through table A3 in (Campbell and Norman, 2012)), the leaf characteristic dimension D ($0.12m$), typical stomatal conductance of lower and upper leaf surface g_{sll} and g_{sul} ($0.2mol/m^2s$), leaf absorptivity α_{leaf} (0.5), relative humidity of the air RH, specific heat of the air C_{pair} ($29.3J/molK$), air pressure P_{air}.

Appendix B

The various models for convective heat transfer coefficients used in the simulations and mentioned in section 3.3 are presented below.

Susorova uses the following equation:

$$h_{vw} = a + bV + cV^2 \tag{B1}$$

Where V is the wind speed and a, b and c are coefficients based on the material roughness. The wall surface is assumed to be of medium roughness, with the coefficients equal to 10.79, 4.192 and 0, respectively.

Morrison & Barfield and Stanghellini calculate h_{vw} based on the Nusselt number, which is given in equations B3 and B4 respectively.

$$h_{vw} = \frac{Nu\lambda}{D} \tag{B2}$$

$$Nu = 0.328Pr^{0.33}Re^{0.5} \tag{B3}$$

$$Nu = 0.37(Gr + 6.92Re^{0.5}) \tag{B4}$$

Where Nu is the Nusselt number, $Re = \frac{DV}{\nu}$ is the Reynolds number, $Gr = \frac{g\beta D^3}{\nu^2}(T_{air} - T_{leaf})$ is the Grashof number. λ is the air thermal conductivity, Pr is the Prandtl number, g is the gravitational acceleration constant ($9.81m/s^2$), β is the thermal expansion coefficient of air ($0.0034K^{-1}$), and ν is the kinematic viscosity of air.

The Ayata and ASHRAE models calculate h_{vw} on the basis of McAdams' equation and are presented in equations B5 and B6, respectively.

$$h_{vw} = 5.9 + 4.1V\frac{511 + 294}{511 + T_{air}} \tag{B5}$$

$$h_{vw} = 5.6 + 4V \tag{B6}$$

Equations B5 and B6 are applicable for $V < 5m/s$, otherwise h_{vw} is calculated from:

$$h_{vw} = 7.2V^{0.78} \tag{B7}$$

Lastly, for Campbell & Norman model, h_{vw} utilizes the boundary layer conductance for heat transfer as:

$$h_{vw} = g_{bh}Cp_{air} \tag{B8}$$

where

$$g_{bh} = 1.4 \cdot 0.135\sqrt{\frac{V}{D}}. \tag{B9}$$

References

Eleftheria Alexandri and Phil Jones. Developing a one-dimensional heat and mass transfer algorithm for describing the effect of green roofs on the built environment: Comparison with experimental results. *Building and Environment*, 42(8):2835–2849, 2007. doi:https://doi.org/10.1016/j.buildenv.2006.07.004.

ASHRAE. American society of heating, refrigerating, and air-conditioning engineers, 2020. URL https://www.ashrae.org/.

Tahir Ayata, Paulo Cesar Tabares-Velasco, and Jelena Srebric. An investigation of sensible heat fluxes at a green roof in a laboratory setup. *Building and Environment*, 46(9):1851–1861, 2011. doi:https://doi.org/10.1016/j.buildenv.2011.03.006.

Ahmet B Besir and Erdem Cuce. Green roofs and facades: A comprehensive review. *Renewable and Sustainable Energy Reviews*, 82:915–939, 2018. doi:https://doi.org/10.1016/j.rser.2017.09.106.

Rosmina A Bustami, Martin Belusko, James Ward, and Simon Beecham. Vertical greenery systems: A systematic review of research trends. *Building and Environment*, 146:226–237, 2018.

Gaylon S Campbell and John Norman. *An introduction to environmental biophysics*. New York, USA, Springer Science & Business Media, 2 edition, 2012.

Sasima Charoenkit and Suthat Yiemwattana. Living walls and their contribution to improved thermal comfort and carbon emission reduction: A review. *Building and environment*, 105:82–94, 2016. doi:https://doi.org/10.1016/j.buildenv.2016.05.031.

F Convertino, G Vox, and E Schettini. Heat transfer mechanisms in vertical green systems and energy balance equations. *Urban Agriculture and City Sustainability*, page 83, 2019a. doi:https:/doi.org/10.2495/DNE-V14-N1-7-18.

Fabiana Convertino, Giuliano Vox, and Evelia Schettini. Convective heat transfer in green façade system. *Biosystems Engineering*, 188:67–81, 2019b. doi:https://doi.org/10.1016/j.biosystemseng.2019.10.006.

KWD Kalani C Dahanayake and Cheuk Lun Chow. Studying the potential of energy saving through vertical greenery systems: Using energyplus simulation program. *Energy and Buildings*, 138:47–59, 2017.

James W Deardorff. Efficient prediction of ground surface temperature and moisture, with inclusion of a layer of vegetation. *Journal of Geophysical Research: Oceans*, 83(C4):1889–1903, 1978. doi:https://doi.org/10.1029/JC083iC04p01889.

Thijs Defraeye, Bert Blocken, and Jan Carmeliet. Convective heat transfer coefficients for exterior building surfaces: Existing correlations and cfd modelling. *Energy Conversion and Management*, 52(1):512–522, 2011. doi:https://doi.org/10.1016/j.enconman.2010.07.026.

R Djedjig, M El Ganaoui, Rafik Belarbi, and R Bennacer. Thermal effects of an innovative green wall on building energy performance. *Mechanics & Industry*, 18(1):104, 2017.

Rabah Djedjig, Emmanuel Bozonnet, and Rafik Belarbi. Analysis of thermal effects of vegetated envelopes: Integration of a validated model in a building energy simulation program. *Energy and buildings*, 86:93–103, 2015.

H Feng and K Hewage. Energy saving performance of green vegetation on leed certified buildings. *Energy and buildings*, 75:281–289, 2014. doi:https://doi.org/10.1016/j.enbuild.2013.10.039.

Yang He, Hang Yu, Akihito Ozaki, Nannan Dong, and Shiling Zheng. Influence of plant and soil layer on energy balance and thermal performance of green roof system. *Energy*, 141: 1285–1299, 2017.

Ghassem Heidarinejad and Arash Esmaili. Numerical simulation of the dual effect of green roof thermal performance. *Energy Conversion and Management*, 106:1418–1425, 2015.

Samy Iousef, Hamid Montazeri, Bert Blocken, and Pieter van Wesemael. Impact of exterior convective heat transfer coefficient models on the energy demand prediction of buildings with different geometry. In *Building Simulation*, volume 12, pages 797–816. Springer, 2019. doi:https://doi.org/10.1007/s12273-019-0531-7.

SEG Jayamaha, NE Wijeysundera, and SK Chou. Measurement of the heat transfer coefficient for walls. *Building and Environment*, 31(5):399–407, 1996. doi:https://doi.org/10.1016/0360-1323(96)00014-5.

Carlos Bartesaghi Koc, Paul Osmond, and Alan Peters. Evaluating the cooling effects of green infrastructure: A systematic review of methods, indicators and data sources. *Solar Energy*, 166:486–508, 2018.

Maria Manso and João Castro-Gomes. Green wall systems: A review of their characteristics. *Renewable and Sustainable Energy Reviews*, 41:863–871, 2015. doi:https://doi.org/10.1016/j.rser.2014.07.203.

John E Morrison Jr and Billy J Barfield. Transpiring artificial leaves. *Agricultural Meteorology*, 24:227–236, 1981. doi:https://doi.org/10.1016/0002-1571(81)90047-9.

Norsk klimaservicesenter. Observasjoner og værstatistikk, 2020. URL https://klimaservicesenter.no/observations/.

Gabriel Pérez, Julià Coma, Ingrid Martorell, and Luisa F Cabeza. Vertical greenery systems (VGS) for energy saving in buildings: A review. *Renewable and Sustainable Energy Reviews*, 39:139–165, 2014.

Babak Raji, Martin J Tenpierik, and Andy van den Dobbelsteen. The impact of greening systems on building energy performance: A literature review. *Renewable and Sustainable Energy Reviews*, 45:610–623, 2015. doi:https://doi.org/10.1016/j.rser.2015.02.011.

Tabassom Safikhani, Aminatuzuhariah Megat Abdullah, Dilshan Remaz Ossen, and Mohammad Baharvand. A review of energy characteristic of vertical greenery systems. *Renewable and Sustainable Energy Reviews*, 40:450–462, 2014. doi:https://doi.org/10.1016/j.rser.2014.07.166.

Shaimaa Seyam. The impact of greenery systems on building energy: Systematic review. *Journal of Building Engineering*, page 100887, 2019.

Cecilia Stanghellini. Transpiration of greenhouse crops: an aid to climate management. 1987.

Tiziana Susca, Stuart R Gaffin, and GR Dell'Osso. Positive effects of vegetation: Urban heat island and green roofs. *Environmental pollution*, 159(8-9):2119–2126, 2011. doi:https://doi.org/10.1016/j.envpol.2011.03.007.

Irina Susorova, Melissa Angulo, Payam Bahrami, and Brent Stephens. A model of vegetated exterior facades for evaluation of wall thermal performance. *Building and Environment*, 67:1–13, 2013. doi:https://doi.org/10.1016/j.buildenv.2013.04.027.

Paulo Cesar Tabares-Velasco and Jelena Srebric. A heat transfer model for assessment of plant based roofing systems in summer conditions. *Building and Environment*, 49:310–323, 2012.

TEK17. Direktoratet byggkvalitet, byggteknisk forskrift (TEK17) med veiledning, 2017. URL `https://dibk.no/byggereglene/byggteknisk-forskrift-tek17/`.

The European Commission's science and knowledge service. Photovoltaic geographical information system, 2019. URL `https://re.jrc.ec.europa.eu/pvg_tools/en/tools.html#TMY`.

UN. 2018 revision of world urbanization prospects. *Population Division of the UN Department of Economic and Social Affairs*, 2018.

Trang Thuy Vo, Alexandru Nichersu, and Jochen Wendel. Modeling, monitoring, and validating green roof and green facade solutions with semantic city models using low cost sensors and open software infrastructures. *Urban Science*, 3(2):39, 2019. doi:https://doi.org/10.3390/urbansci3020039.

YAGA AS. Yaga dugpunkskalkulator, 2020. URL `https://www.yaga.no/dpkalkulator.php`.

Dynamic modeling of heat pumps for ancillary services in local district heating concepts

Mads Pagh Nielsen   Kim Sørensen

Department of Energy Technology, Aalborg University, Denmark, {mpn,kso}@et.aau.dk

Abstract

This paper aims to describe the steady-state and dynamic heat pump models developed to study their abilities in ancillary services as well as the inter-connection between these, electrical boilers and thermal storage with the aim to balance power and heat production for a given case.

An hourly steady-state system model was developed to understand the overall operational characteristics of the system for a given heat demand case in the area Kolt-Hasselager-Ormslev near Aarhus in Denmark. The model showed an annual average COP above 3,5 for a serial connected heat pump system.

Detailed thermal and dynamical models of the heat pump system were developed. The models show that it will be possible to use heat pumps successfully in ancillary services. The turn-up is unproblematic but the turn down of the heat pump will be limited in a non-liquid overfed system due to risk of liquid formation in the evaporator, requiring additional heating.

Keywords: heat pumps, dynamic modeling, district heating, thermal properties, ancillary services.

1   Introduction

The Danish heating sector has successfully evolved over the past 50 years. Today, more than 60 % of the domestic heating in Denmark is supplied by district heating. Numerous developments like urbanization, city densification, reduction in heating consumption through better insulation standards etc. challenges the efficient operation of the district heating systems.

Simultaneously, the development of wind farms has created a surplus of electricity in Denmark requiring power balancing. Heat pumps coupled with electrical boilers and heat storage could provide such balancing maintaining an efficient heat supply system. For automatic frequency response reserve actions the response time and operation of the heat pump should meet certain standards (max. 150 s up-/down regulation for Frequency Containment Services) depending on the way the heat pump acts as backup for power and net frequency fluctuations. The downward and upward regulation compensation could potentially greatly benefit the economy of heat pump systems if the dynamic response is fast (Energidataservice, 2020).

The "Local Heating Concepts" ForskEL project funded by the Danish Government has through detailed inputs from several asset owners investigated the possibilities of integrating localized heat pumps in district heating and power grids. Through laboratory emulations at Aalborg University, Department of Energy Technology, the models will later be coupled with real power system components enabling a realistic dynamic emulation.

There are not many works published on the operation of large-scale heat pumps for ancillary services in Denmark. However, recently, (Meesenburg et al., 2020) published a paper on the optimization of heat pumps for this purpose.

Detailed heat pump models designed for both parallel and serial coupling have been developed and simulations have been carried out. As a case, the district heating grid at Kolt-Hasselager-Ormslev near Aarhus has been chosen. The system operation and some system configurations have been investigated.
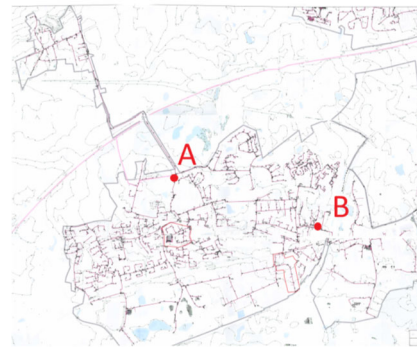
Figure 1. Kolt-Hasselager Ormslev district heating grid.

Currently, the district heating area is supplied by a large-scale plate heat exchanger linked to the central CHP-plant in Aarhus and industrial heat suppliers (A) and three oil-fired backup boilers (B). In the coming years the district heating demand will increase significantly requiring additional capacity. The supplier would also like to replace the oil-fired boilers with heat pumps.

2   Traditional operational modeling

Degree day corrected heat demand data for 2017 in the area is shown in Figure 2 along with the heat rate demand curve. The heat demand has furthermore been corrected to reflect the expected future heat demand as predicted by the heat supplier AVA (AffaldVarme Aarhus). The area around Ormslev shown in the upper left corner of Figure 1 is expected to be expanded significantly with several new areas with family dwellings thus requiring a total of around 5-10 MW additional heat.
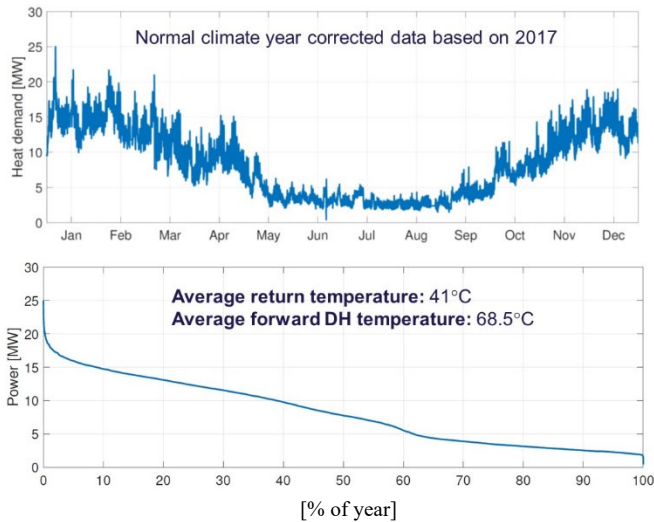
Figure 2. Corrected demand and duration curve.

The annual average forward and return temperatures of the district heating water are also shown in Figure 2. The forward temperature typically varies in the interval from 65-75 °C (in rare cases up to 80 °C) and the return temperature varies from 36-45 °C.

Based on the local conditions in the area, it was decided that a 12-16 MW$_{heat}$ maximum heat output heat pump consisting of six to eight 2 MW$_{heat}$ heat pumps coupled in series and a 3-5 MW electrical boiler along with a 200 MWh heat storage would be suitable to cover the load of the area. The supply system will also in the future be coupled to the centralized district-heating grid in Aarhus, which covers the peak demands.

For comparison, a conventional logical operational scheme disregarding ancillary service operation was found by establishing optimum threshold max and min limiting costs for the heat pump operation. This was established solving an MILP-problem minimizing the annual cost of system operation subject to constraining the heat demand and capacities of the individual units. The logical operational scheme is show in Figure 3.
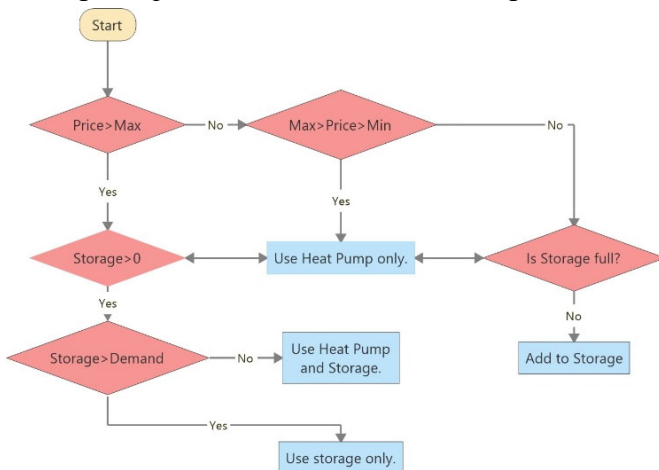
Figure 3. Logical conventional operational scheme.

Modeling a year on an hourly basis results in the following operation of the different assets shown in the following figures. Figure 4 shows the operation of the heat pumps. Each heat pump is in the modeling considered to operate by on/off control.
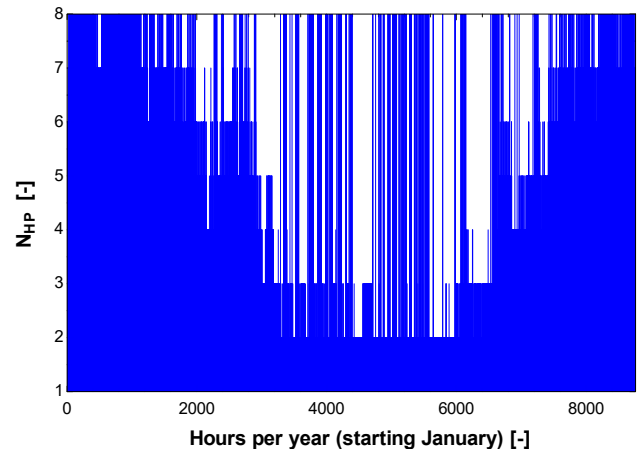
Figure 4. No. of heat pumps on/off each hour for a year.

Figure 5 shows the operation of the electrical boiler.
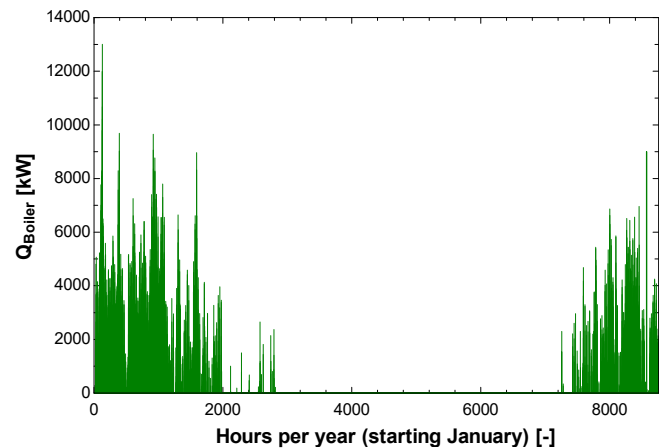
Figure 5. Operation of the electrical boiler.

It is clear the electrical boiler with 2017-prices mainly operates during the cold months in case of peak demands. The current taxation on electricity in Denmark will however in near future be changed, where the tax on using electricity for heating purposes will be lowered from 155 kr./MWh to 4 kr./MWh. This will strongly further favor the operation of electrical heating and heat pumps in the future.

Figure 6 illustrates the operation of the storage. A 200 MWh stratified hot water storage @90 °C corresponds to about 3000 m³ of water, which is not a very huge thermal storage considering the system size. The reason for this is the lack of space in the district heating area. There are only very few available locations for a large storage tank.
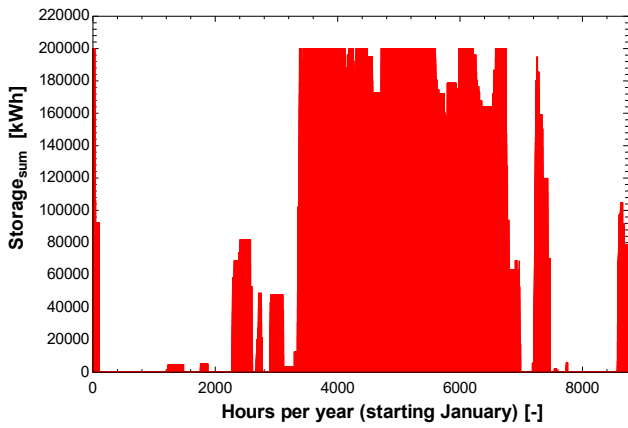
Figure 6. Operation of the storage tank.

Heat is, using the current operational scheme, typically accumulated during the Summer period. This operational pattern will clearly change significantly in case of ancillary service operations where the pattern is expected to be much more dynamic over the year. It was assumed that the storage stratification is kept at all times in this overall model. In the more dynamic case this is likely not to be true in some situations and thus a detailed model of the stratification will be implemented in the dynamic modeling. The average annual heat pump system COP was 3,52 and the overall annual cost of supplying heat for the area is 12,8 million Danish kr.

3 Mathematical modeling

An ammonia-based system with flooded evaporator and two-stage screw compression was used as base case for the heat pump modeling. The process is illustrated in the PI-diagram in Figure 7 and the log(p),h-diagram in Figure 8:
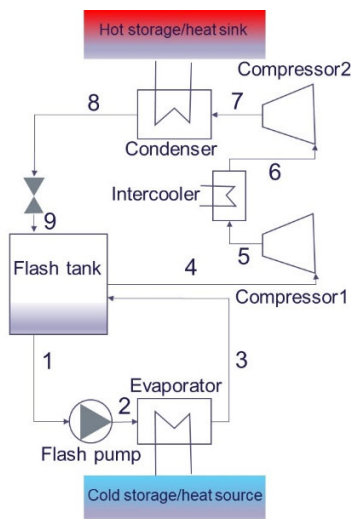
Figure 7. PI-diagram of a single heat pump.

The system can optionally be coupled with either a heat source or a "cold" thermal storage in the source-side. This storage could also receive low grade heat if other sources are available. On the sink side, normally the heat pump is coupled to a thermal storage (hot storage).
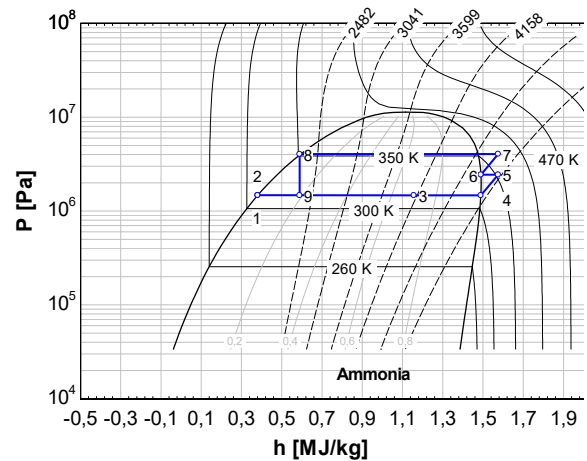
Figure 8. Log(p),h-diagram of the heat pump process.

The steady-state models have been developed in EES (Engineering Equation Solver) which is an advanced numerical non-linear equation solver with built-in thermal property functions and the dynamic heat pump modeling was conducted using MATLAB. The steady-state models are based on the solution of non-linear equation sets coupled with thermal property models. The dynamic MATLAB model is based on the solution of a set of differential algebraic equations (DAE's). The thermal storage has been modeled one-dimensionally to account for basic stratification constraints. A discretized DAE-model was used to enable a stabile model in both filling and storage mode – see (Bitzer et al., 2008).

3.1 Thermal /calorimetric property routine

A simple routine using the Peng-Robinson cubic equation of state to determine thermal and calorimetric properties for the model was developed to have a fast property library for real-time emulation. The model must clearly be able to simulate faster than real time. The overall solution principle is shown in Figure 9.
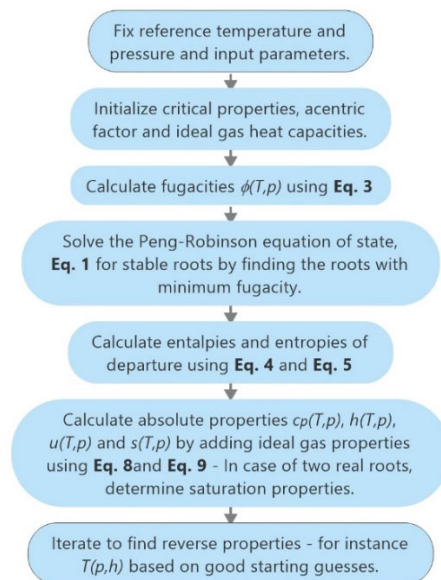
Figure 9. Use of the Peng-Robinson cubic equation of state to estimate thermal and calorimetric properties.

Due to large inaccuracies in the liquid region for the polar fluids water and ammonia, simple regressions based on property data from EES were used in these cases. The methodology was compared to accurate data from EES based on the reduced Helmholtz equation of state for water and ammonia and for the relevant temperatures and pressures for heat pump operation, the maximum deviations was found to be less than 7 % (worst case) and generally significantly lower (typically 2-3 %). This is considered acceptable for the fast dynamic model.

The generalized cubic equation of state in terms of the compressibility factor, Z [-], to be solved for its roots can be written as (Matsoukas, 2012)

$$Z^3 - (B' + 1)Z^2 + A'Z - A'B' = 0 \qquad (1)$$

where the coefficients B' [-] and A' [-] defined as

$$A' = \frac{aP}{(RT)^2} \ \& \ B' = \frac{bP}{RT} \ \text{w.} \ a = \frac{27}{64}\frac{(RT_c)^2}{P_c} \ \& \ b = \frac{1}{8}\frac{RT_c}{P_c} \quad (2)$$

where P [Pa] is the total pressure, R [J/(mol*K)] is the universal gas constant, T [K] is the temperature, T_c [K] is the critical temperature and P_c [Pa] is the critical pressure. The constants a [J*m³/mol²] and b [m³/mol] are the constants of the state equation. The fugacity coefficient ϕ [-] was found as:

$$\phi = e^{\left(Z - 1 - \ln(Z - B''') - \frac{A'}{2\sqrt{2}B'}\ln\frac{Z+(1+\sqrt{2})B'}{Z+(1-\sqrt{2})B'}\right)} \quad (3)$$

The molar specific departure enthalpies h^D [J/mol] and entropies s^D [J/mol] were found analytically for the Peng-Robinson equation using eq. 4-7

$$h^D = RT(Z - 1) + \frac{T\left(\frac{da}{dT}\right) - a}{2\sqrt{2}b} \ln\frac{(1+\sqrt{2})B'+Z}{(1-\sqrt{2})B'+Z} \quad (4)$$

$$s^D = R\ln(Z - B') + \frac{\left(\frac{da}{dT}\right)}{2\sqrt{2}b} \ln\frac{(1-\sqrt{2})B'+Z}{(1+\sqrt{2})B'+Z} \quad (5)$$

where the derivative $\frac{da}{dT}$ [J*m³/(mol²*K)] is found as:

$$\frac{da}{dT} = -0{,}45724 \frac{R^2 T_c \left(1 + \Omega(1 - \sqrt{T_r})\right)\Omega}{P_c\sqrt{T_r}} \quad (6)$$

and Ω is

$$\Omega = 0{,}37464 + 1{,}54226\omega - 0{,}26992\omega^2 \quad (7)$$

where ω [-] is the acentric factor.

Some corrections has been proposed to particularly (7) to obtain more accurate results. For instance (Strykjek and Vera, 1986). However, the found accuracy is acceptable for this study and the corrections would

lower calculation speed which is not desirable in this application.

Specific enthalpies and entropies in [J/kg] are then found using the reference values, the departure enthalpies and the ideal gas enthalpies as:

$$h = \frac{h^D - h^{reference} + h^{idealgas}}{M_{fluid}} \quad (8)$$

$$s = \frac{s^D - s^{reference} + s^{idealgas}}{M_{fluid}} \quad (9)$$

M_{fluid} [kg/mol] is the molar mass of the given fluid. The ideal gas enthalpies were found based on analytical integration of polynomials of the ideal gas specific heat capacities.

The Matlab-function *fzero* was used to find reverse properties using appropriate starting guesses and a minimization algorithm. Guesses were defined so that the algorithm converged in all cases.

3.2 Steady-state model of the heat pump

A steady-state model based on power and mass flow balances was setup for the system in EES. The following operational assumptions were used in the modeling of the two stage liquid overfeed ammonia based heat pump cycle:

- The district heating water was assumed to have a forward absolute pressure of 3 bar and a return pressure of 2 bar.
- Volumetric and isentropic efficiencies were variable and found using the SABROE COMP1 compressor data tool (Sabroe, 2020). For the 2MW system nominal value are approximately η_s=0,8 and η_v=0,9. The flash tank pump was assumed having an isentropic efficiency of η_p=0,8.
- Condenser and evaporator minimum approach temperatures were fixed to 3 K.
- Pressure drops in both the condenser and evaporator were assumed to be 3 kPa.
- In the cycle superheating was assumed to be 1 K and sub-cooling was 5 K.
- The compressor displacement was 0,4428 m³/s.
- The intermediate pressure between the compression stages was found optimizing the COP.
- The quality at which refrigerant leaves the liquid overfeed evaporator was fixed to x=0,7

The model was developed using EES's "MODULE" functionality making it easy to couple heat pumps in series and parallel on both the source and the sink side of the heat pumps. An example of serial coupling 2 heat pumps on the source and sink side, shown in Figure 10.
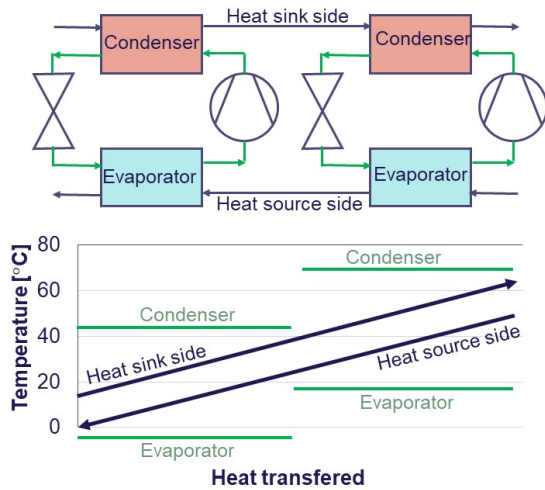
Figure 10. Serial coupled heat pump on sink/source side.

As an example, a system with 5 similar serially coupled heat pumps in terms of both condenser and evaporator having evaporation temperatures of 3; 7; 11; 15 and 19 °C and condensing temperatures of 48; 54; 60; 66 and 72 °C would lead to an overall system COP of 4,46 where a single stage reference system with the same minimum and maximum temperatures would only have a COP of 3,53. The investment cost in the more complex system would of course be significantly higher, so the choice of system topology is to a large extent a choice based on the operational characteristics and possibilities. A multi-stage system would have the advantage, that the heat pumps can rapidly be decoupled which would give a rapid down-regulation response.

The heat exchangers were modeled using a basic LMTD-algoritm. A function was used to avoid numerical problems with the LMTD-formula during iterations in EES as follows (Herold et al., 2016). The returned value LMTD is the logarithmic mean temperature difference, which is a function of the four terminal temperatures T1 to T4.

```
FUNCTION LMTD(T1;T2;T3;T4)
  dTa=T1-T2
  dTb=T3-T4
  IF (dTa=dTb) THEN
    LMTD=dTa {Test for singularity in LN function}
  ELSE
    IF (dTa < 0)  or (dTb < 0) THEN {Test for impossible
intermediate heat exchanger solution and force LMTD to nearly
zero if so}
      LMTD=1e-9
    ELSE
      LMTD=(dTa-dTb)/LN(dTa/dTb)
    ENDIF
  ENDIF
END
```

The steady-state model was used to provide inputs to the dynamic model. Thus, a number of multi-dimensional regressions were made using the generalized least squares algorithm in EES to the cycle temperatures and flows as function of the sink and source temperatures.

The regressions were compared with the rigorous model to evaluate their accuracy as shown in Figure 11.
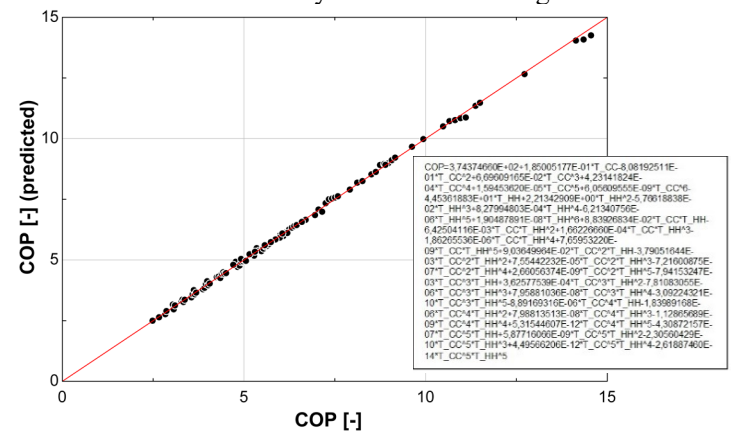
Figure 11. Example on regression comparison.

3.3 Dynamic heat pump modeling

The dynamic heat pump model is fundamentally based on power and mass flow balances using the dynamic forms of the first law of thermodynamics and the continuity equation. The general thermal properties are determined using the previously described steady-state model assuming relatively fast dynamics in terms of changing pressure in the system. The evaporator and condenser are considered as the primary dynamic elements, these were divided into sub-heatexchangers to model the phase change the ε-NTU method was used to model each of the sub-heatexchangers - see Figure 12.
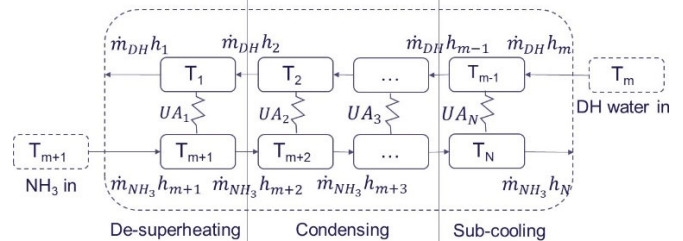
Figure 12. Discretization of latent heat exchangers.

Where N [-] is the number of discretized nodes in the heat exchanger. A grid independency study showed that a value of N=50 (i.e. 25 points on each side of the heat exchanger) would be reasonable for the real-time model. As illustrated, the latent heat exchangers are further divided into respectively de-superheating, evaporation and sub-cooling for the condenser and in evaporation and superheating for the evaporator.

The overall governing equations are in their discretized form in the flow direction (x) are formulated for the condenser for respectively the district heating water side (*DH*) (10) and the refrigerant side (11)

$$V_{DH}\rho_i \frac{du_i}{dt} = \rho_i A_{DH} v_i (h_{i+1} - h_i) + UA_i(T_{i+m+1} - T_i) \quad (10)$$

$$V_{NH_3}\rho_{i+m+1} \frac{du_{i+m+1}}{dt} = \rho_{i+m+1} A_{NH_3} v_{i+m+1}(h_{i+m} - h_{i+m+1}) + UA_i(T_i - T_{i+m+1}) \quad (11)$$

Here *V [m³]* is the volume of each element, *A [m²]* is the cross sectional area at the element, v [m/s] is the fluid

velocity, h *[J/kg]* is the specific enthalpy, *UA [W/K]* is the total heat transmission coefficient multiplied by the element heat transfer area and *T [K]* is the temperature. The energy balance is formulated in terms of internal energy so the equations are valid for all three heat exchanger sections. The mass balance becomes:

$$\frac{dm}{dt} = \rho(x,t)_{in}v(x,t)_{in}A - \rho(x,t)_{out}v(x,t)_{out}A \quad (12)$$

Energy and mass balances for the evaporator were formulated almost exactly identically as shown in (10)-(12).

In each element of the discretized heat exchangers, the total heat transmission coefficient was calculated. The Dobson and Chato relation for flow condensation, as described in (Nellis and Klein, 2009) for condensation over horizontal tubes was used to estimate the convective heat transmission coefficient for the ammonia in the condensing part. For the evaporator the Shah correlations for flow boiling were used as described in (Nellis and Klein, 2009). For the remaining single-phase pipe flow regular correlations for friction factors and Nusselt numbers for developing flow in pipes considering whether the flow is laminar or turbulent was used as described in section 5.2.3. and 5.2.4 of (Nellis and Klein, 2009). It was assumed that all heat exchangers were of the shell and tube type and the heat exchanger dimensions were defined following the procedure described in (Stephan et al. 2010). The design was done considering heat exchanger pressure losses.

3.4 Solution of governing DAE set

The thermal relations and the sets of ordinary differential equations from the discretized heat exchangers forms a differential algebraic equation set (DAE). All algebraic equations in the DAE set are time-independent, thus the system is of index 1. The ODE15s solver in Matlab was used to solve the system of equations. Using this solver, all algebraic equations must be rewritten on the form of pseudo differential equations and a mass matrix **M** is defined with binary values of either 1 for differential equations and 0 for algebraic equations in the equation set. The mass matrix is practically implemented by multiplying it to the differentials as shown in (13).

$$\boldsymbol{M}(t,y)\frac{dX}{dt} = f(t,y) \quad (13)$$

The solver convergence criterions was set so the relative tolerance was 10^{-3} and the absolute tolerance 10^{-6}.

4 Results

Results of the retrofitting of the system topology and operational optimization of the system considering steady-state operation have been investigated.

Furthermore, the dynamic capabilities and thermal and practical response time limitations in the heat exchangers of the heat pump was investigated considering both ramp-up-, ramp-down- and the response in terms of power and heat. A theoretical start-up where the heating capacity is given in terms of the relative district heating water temperature is given in Figure 13 (72 °C DH water@100 %):
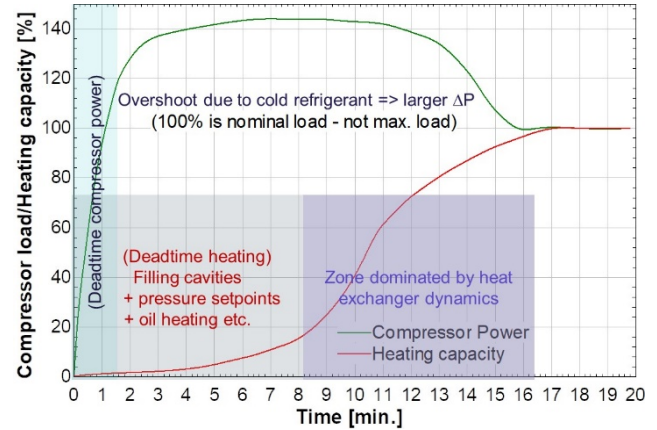
Figure 13. Typical cold-start response of the heat pump.

Nominal load refers to a heat output of 2 MW from the condenser. Max. load is the start-up capacity but at temperatures unusable for district heating. It can be seen that the start-up response in terms of heating is significantly longer than the requirements for frequency services on the power grid. Thus, operation of the heat pumps in a partial load-condition or a forecast-based scheduling would be required to meet the heat demand. In reality, the heat pumps are never operated from a cold conditions. There are typically oil separators and heat tracers installed on the compressor suction line [Johnson, 2020]. The ramping up and down of the compressor power can be significantly faster if the systems remains pre-heated or in a partial-load, so the requirements in terms of power uptake/outage can be fulfilled before steady-state is reached. The measured load-torque versus the induction motor drive torque is seen for a start-up of a screw compressor based 500 kW heat pump cycle in Figure 14 [Johnson, 2015].
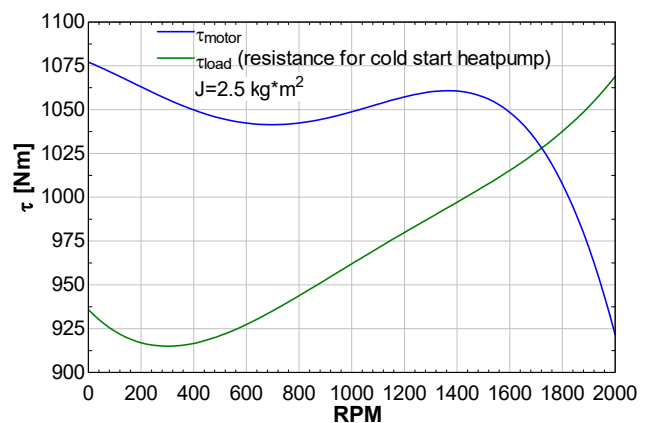
Figure 14. Load & induction motor torque during startup.

Solving the initial value problem shown in (14), the compressor approach to steady-state relative to the compressor RPM can be found as shown in Figure 15.

$$J_{compressor} \frac{d\omega_{compressor}}{dt} = \tau_{induction\ motor} - \tau_{load,HP} \tag{14}$$
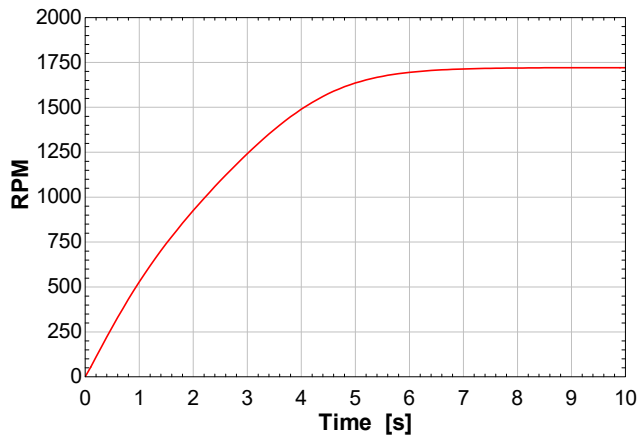
Figure 15. Approach to steady-state compressor RPM.

It can be concluded from the plot that the compressor /compressors even in worst-case does not pose a major limitation to the transient operation of a heat pump.

However, ramping down, the evaporator and condenser conditions will restrict the operation. Typical temperature curves in the condenser at steady-state are shown in Figure 16:
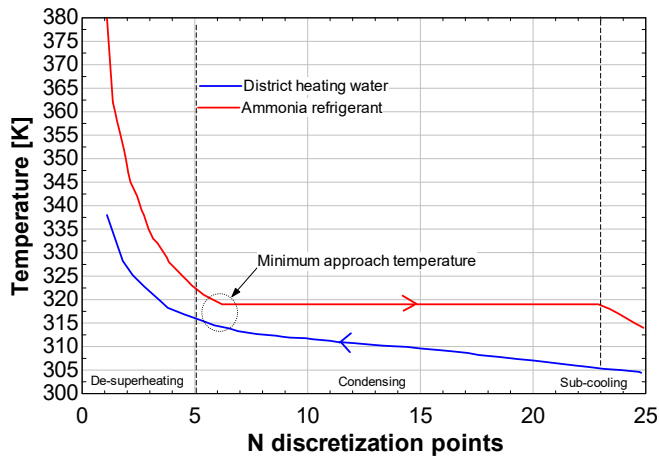
Figure 16. Condenser temperature curves.

Currently, the model has no active control scheme built-in. Thus the model is only controlled based on the initial conditions for each modeling run. Thus parameters like superheating and sub-cooling are difficult to control accurately. In future versions, control of the compressor, the pump and the valve should be implemented to address this.

The progression of the district temperature over time going from a stand-by heated condition to nearly full load (i.e. the zone dominated by the heat exchanger dynamics) is seen in Figure 17. As mentioned, the lubrication and cooling system is preheated at all times. A real cold start thus never really exist however, the heat exchangers could be cooled significantly if the system is left out of operation for longer periods during winter.

The modeling run corresponding to Figure 17 was using the following initial system conditions:

- Initial mass flow of refrigerant: 0 kg/s.
- Mass flow of liquid heat source: 90 kg/s.
- Mass flow of the district heating water: 15 kg/s.
- Initial district heating water temperature: 40 °C.
- Temperature of the heat source (from a cold storage tank): 15°C.
- Initial refrigerant temperature in the cycle: 8°C.
- Initial refrigerant pressure in cycle: 1100 kPa.
- Initial temperature of hot source in: 15 °C.
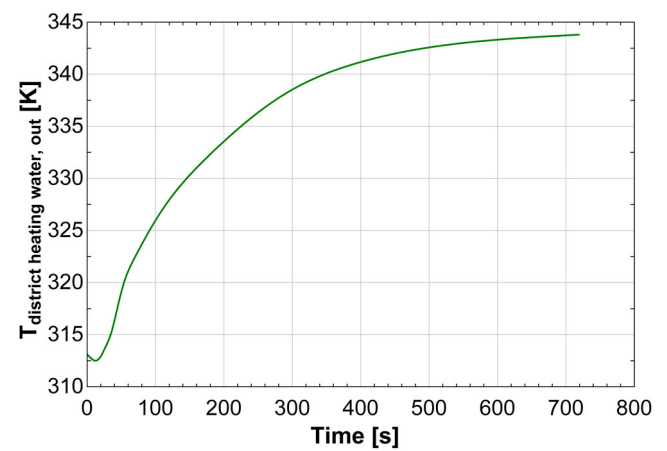
Figure 17. Dynamic response of district heating water temperature going from a "cold" condition.

It can be seen in Figure 17, that response has some dead-time in the beginning. When designing actual controllers, it is important to compensate for this. Solving the model, it was found that it in most cases solve significantly faster than real time on an ordinary Surface 6 Pro laptop computer. In rare cases, the model was slightly slower than real time but it is expected that the multicore PC's attached to the real-time emulation system at Aalborg University, Department of Energy Technology will be able to simulate significantly faster. Coupling the model with the stratified heat storage model however currently still give some challenges. The simulation times in this case increase significantly. A preliminary model simulating 48 hours of system operation including the storage has proven to be less than half as fast as real time. The models have to be optimized to address these challenges.

4.1 Discussion of liquid overfed operation

The liquid overfed system is often an economic tradeoff between a better heat transfer per unit volume evaporator and the cost of additional pump, piping and the flash tank. When operating the heat pump for ancillary services, there are however clear advantages of

the liquid overfed system. When ramping down, it is important to avoid condensing of the refrigerant in the evaporator in case the system is not operated as liquid overfed. This will limit the time required to go to a partial load or stand-by situation of the heat pump. Introducing a simple energy balance for the pipe-material as shown in (15), we can assume that the saturation temperature will change rapidly with the thermal conditions of the refrigerant during ramping of the mass flow from 100 % to 50 % load. The temperature of the refrigerant in the evaporator will change almost instantly with the mass flow.

$$\frac{du_i}{dt} = \frac{\dot{m}(h_i - h_{i+1}) + \dot{Q}_{suction\,pipe}}{\rho_{pipe}\Delta V_{pipe}} \qquad (15)$$

The suction pipe thickness was assumed to be 6 mm thick, it is assumed adiabatic and have a specific heat capacity similar to that of stainless steel since this will be a likely material working with ammonia as refrigerant. In this calculation, it is assumed that the refrigerant leaves the evaporator with a certain degree of superheating to avoid liquid in the compressor. Figure 18 shows the results on the suction pipe temperature when ramping the heat pump linearly from 100 % to 50% mass flow over the required 150 s for Frequency Containment Processes (FCR).
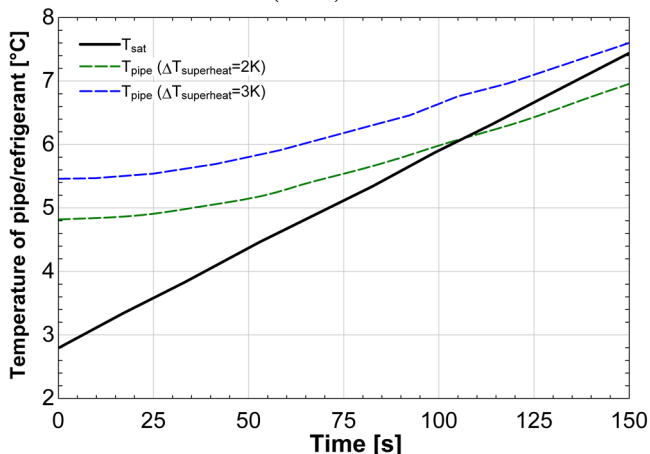
Figure 18. Ramping from 100-50% mass flow in 150 s.

Two temperature trajectories are shown. One with a superheating of 2 K and one with 3 K. These can be compared to the thick black line showing the saturation temperature of the refrigerant. The larger the superheating, the lower the cycle COP will be.

Many commercial systems are constructed with insulated suction pipes and superheating, so this could be a challenge in these cases. The problem can be resolved in multiple ways; either by increasing superheating (i.e. accepting lower COP), implementing electrical heating or constructing/ramping the system in a different manner avoiding condensation.

5   Conclusion

A fast numerical model of a large-scale heat pump system, which can simulate faster than real time so it can be included in the Smart Energy real-time emulation lab

at Aalborg University has been developed. Different heat pump topologies have been considered in the paper and a rigorous, scalable and modularized thermal model was developed. A fundamental dynamic heat pump model has been successfully developed with focus on the detailed modeling of the latent heat exchangers. However, the model is yet to be integrated with the thermal storage model and other heat producing assets and heat distribution networks in the system and implemented in the real-time lab. Here more work still remains to have a useful modeling framework capably of solving in real time. The economic aspects of coupling heat pumps in series must also be addressed. The same goes for the considerations on a liquid overfed system. The system offer many advantages but could have a larger investment cost. Operating without a flooded evaporator however introduces challenges when ramping down the heat pump to partial load. Initial estimates indicate a undesired superheating of nearly 3 K in the first scompression stage suction line is required to avoid liquid formation in the evaporator in a non-liquid overfed system. More work looking into possible solutions and their consequences must be conducted.

References

M. Bitzer, T. Kreuzinger, and W. Marquardt. Mathematical modelling of a domestic heating system with stratified storage tank. *Mathematical and Computer Modelling of Dynamical Systems*, 14 (3): 231-248, 2008.

Energidataservice.dk (homepage). Page with historical data on the cost ancillary energy services. Assessed 3/7 2020.

K. E Herold, R. Radermacher, and S. A. Klein. *Absorption Chillers and Heat Pumps*. CRC Press. 2016.

Johnson Controls. *Engineering, Service and Maintenance Manual*. SMC/TSMC and HPC 100 Mk4. 2015.

Johnson Controls. Personal communication with employees at Johnson Controls, 2020.

T. Matsoukas. *Fundamentals of Chemical Engineering Thermodynamics*. Prentice Hall International Series. 2013.

W. Meesenburg, W.Markussen, T. Ommen and B. Elmegaard. Optimizing control of two-stage ammonia heat pump for fast regulation of power uptake. *Applied Energy*. 271. 2020.

G. Nellis and S. Klein. *Heat Transfer*. Cambridge University Press, 2009.

Sabroe Compressor data tool COMP1. https://www.sabroe-.com/en/products/sales-tools/comp1. Assessed 2/6 2020.

P. Stephan (editor). *VDI Heat Atlas 2nd edition*. Springer-Verlag, Berlin, Heidelberg 2010.

R. Stryjek and J. H. Vera. PRSV: An improved Peng-Robinson equation of state for pure compounds and mixtures. *The Canadian Journal of Chemical Engineering*. 64 (2): 323-333. 1986.

Multi-Scale Smart Building Simulation Platform for Energy, Market and Automation Solutions

Jean-Nicolas Louis   Jari Pulkkinen

Water, Energy and Environmental Engineering, University of Oulu, Finland,
`jean-nicolas.louis@oulu.fi`, `jari.pulkkinen@oulu.fi`

Abstract

The increasing presence of smartness in buildings to increase the energy efficiency of these buildings makes it difficult to properly anticipate the impact of these technologies. The development of digital tool for creating residential load profile are prolific but their use, especially applied to smart buildings being able to combine both power and thermal energy management, are unavailable. This research presents a tool developed to integrate power and heat demand using a flexible topology to define buildings energy needs across time resolution. It also integrates small-scale renewable energies such as wind and solar to increase the production of local electricity. The entire model can be parametrised and personalised through its user interface. The model demonstrated robust simulation load profiles generation throughout all time resolutions

Keywords: smart buildings, high resolution, heat and power, demand response, real-time pricing

1 Smart buildings

The use of smart buildings and its role in shaping the new energy efficiency plans from the EU Green deal is crucial (EU Commission, 2019). The smart building is becoming a new standard in both the standardisation body and in the regulation with the implementation of the Smart Readiness Indicator (SRI) in EU. This comes in combination with the smart grid standards that are becoming more and more implemented in real-life microgrids, or in the establishment of positive energy district such as those built in frame of the EU lighthouse project Making-City. International research activities has put a strong focus on building flexibility and their role in integrating more distributed variable renewable energies onto the distribution network. Such research are built through the International Energy Agency (IEA) and its series of annex targeted on Energy in Buildings and Communities program (IEA-EBC) in its Annex67 (Jensen et al., 2017) and its future Annex82 on Energy Flexible Buildings.

Multi-energy vector (Electricity, heat, hydrogen, etc..(Orecchini and Santiangeli, 2011)) applied to building simulation and optimisation have been deployed for at least 20 years with for instance the doctoral thesis of Stokes (2005) on producing fine-grained load model to support low voltage power network. These models produce load profiles for network performance analysis but are usually not publicly available, nor the code that goes along with the models. In more recent years, tool for simulating multi-scale simulation with medium horizons were developed and integrated Stokes' work into an open-source Excel VBA tool (McKenna and Thomson, 2016). Little by little, models are becoming more robust and integrate occupancy from statistics (Wang et al., 2018) or from appliance usage probability (Yilmaz et al., 2017). There is however a need for a new tool that combines latest development on the field of flexible energy and that is open source and accessible by all.

The **S**mart **Bu**ilding **M**odel (SBuM) (Louis et al. 2016) was originally created to measure the impact of demand response and occupant behaviour on the building's energy performances using different levels of control. Controls included basic information-based system to raise awareness to full control over the appliances installed in the building. It has further been developed to integrate the thermal performance of buildings and integrate electric heating technologies as a demand response tool for smart building (Pulkkinen and Louis, 2019).

This work aims at developing an open and user-friendly tool to setup multi-scale building energy models, integrating variable time dependent demand response programmes and providing detailed load profiles that are representative of input profiles. The model shall be adaptable for multi-regional study and work under various weather conditions. The model outputs should cover detailed energy analysis, monetary impacts of different demand response rate, and account for the environmental impact of the related energy consumption.

2 Heat and power model

The SBuM model is a Markov-chain model developed using modules that are plugged in into the main model. It is composed of multiple modules that integrates the following: a small-scale energy production, control options for appliances and heating systems, and a home appliance model. Other modules are being developed to cover the full range of demand response possibility for smart building such as the domestic hot water. Unlike most mathematical model presented in the literature, the SBuM model is available open source (`https://github.com/jeanlouisnico/SBuM`) and can be run through

the MatLab™ simulation platform.
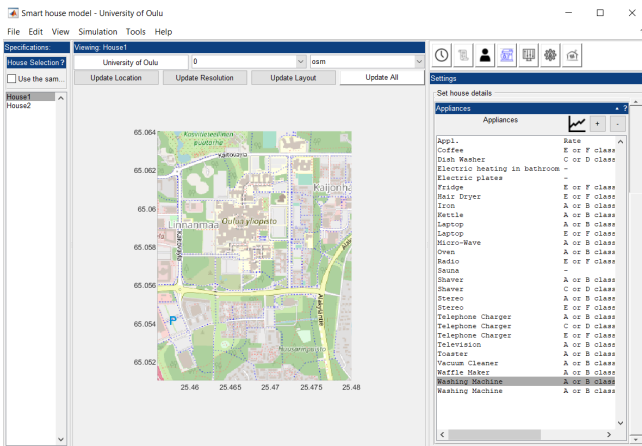
Figure 1. Smart Building Model (SBuM) interface seen from the MatLab™ simulation platform

Interface allows to have flexible and easy-to-use tuning of the building and all the variables considered in the SBuM tool are tunable. It allows as well to do reporting and retrieve all building characqteristics that are used as input variables to the model.

2.1 Input data

A set of input data are required to be able to run the model and use its possibilities. There are six (6) type of input data category needed to fill in and are visible on the right-hand side of Figure 2: the weather files, the household description from appliances to users, the time resolution and horizon of the simulation, the optional power productions units, the type of electricity contracts used, and the building characteristics detailing its physical properties (see section 2.2.2).
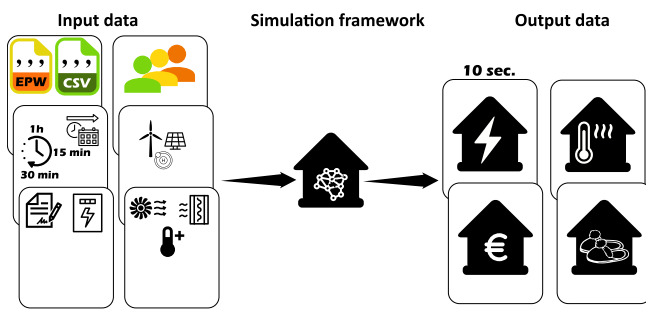
Figure 2. SBuM architecture from input data, modelling, and output variables.

2.1.1 Climatic data

The model is flexible on the type of weather format to be included and any EPW weather file can be given as an input to the model (Crawley et al., 2001). EPW files are available in open access[1] and may be modified to represent

---

[1] http://climate.onebuilding.org

new weather for extreme conditions or represent future climatic conditions considering the different RCP scenarios and other General Circulation Model (GCM). Most of the data are gathered with an hourly time resolution, different weather data were linearly interpolated to match the simulation time-frame e.g. outdoor temperature, with some randomness in the interpolation e.g. solar radiation and wind speed.

2.1.2 Households characteristics

The SBuM model includes a set of variables that define the house to be modelled. All appliances that have a power signature may be included in the model and are characterised by their daily load profile probabilities, usage duration, weekly maximum usages, nominal power ratings, weekday/weekend distributions and control options for demand-side management implementation. The appliances characteristics and usage statistics are calibrated using the EuP reports, specifically adapted for Finland and the Finnish national statistics on energy use; they are however tunable to match any other performances when needed or reflect updated information compared to the previous EuP studies. For detailed information on households' specifics, refer to Louis et al. (2016). Similarly to Yilmaz et al. (2017), the occupancy distribution is deduced from appliances activities that are set in the beginning of the simulation. Their daily distributions are alterable to model any appliance distribution that the model needs to consider during the simulation.

2.1.3 Small-scale production system

Small-scale production units that can be utilised in the model include so far only solar photovoltaic panels, but experimental modules for wind turbines and fuel cells (Proton Exchange Membrane model) are also available. These models are somewhat simplified as multiple variables, such as the terrain and relief of the surroundings are not used in the model. The PV-model requires a set of 12 inputs characterising the photovoltaic modules. The methodology and the used equations are established from Luque and Hegedus (2010) and thus the made assumptions come from the book. The methodology consists calculating the voltage and current of one single cell of a solar module and then, depending on the solar irradiance, evaluating the power output of a series of cells. It is possible to outline real performance from technology available on the market.

2.1.4 Simulation time resolution and horizon

The model works for different time scale ranging from 1 minute to 1 hour time step with a time horizon hypothetically open, and is linked to the exogenous information input into the model - essentially the weather file. Input data regarding appliance usage is provided on an hourly basis and therefore down-scaling is necessary. This part is detailed in section 3.

2.1.5 Electricity contracts

Demand response programmes including real-time pricing, Time of Use tariffs, direct load control and other as listed in the literature (Pallonetto et al., 2020) are included in the model with the defined simulation time resolution. The dynamic pricing can further be limited upwards and downwards if necessary.

2.2 Modelling framework

The SBuM model is built into sub-modules that includes the power production, power consumption, heat demand, and control systems. The following sections briefly introduce the three main elements of the SBuM.

2.2.1 Power consumption

Power consumption is composed of 4 variables: electricity consumption for appliances, ventilation, heating/cooling elements, and domestic hot water. Power consumption distribution from appliances is described by Louis et al. (2016) and covers the fields of appliance modelling, user type considerations, demand response potential for appliances, and occupancy scenarios. Occupancy is deduced from the power consumption profiles from the appliances and not from occupancy profiles used as input to the model. This is to ensure that the power consumption from appliances and the active occupancy scenario are dependent and therefore coherent. Further on, the thermal needs of buildings were characterised to include a number of new variables that define the buildings physics and the ventilation technology used (Pulkkinen and Louis, 2019). Appliance specifics are given default values taken from the Remodece project (Almeida and Fonseca, 2006), but includes also other sources like the monitoring and research work from Murray et al. (2016, 2017) or Reinhardt et al. (2012)[2]. The power module is and the occupancy is further used by the thermal demand model.

2.2.2 Thermal demand

The thermal demand of the model includes the thermal needs of the building and use thermal mass as a thermal storage to optimise the use of electricity when electric heating is used. The model is a white-box model, implemented by using the National Building Codes of Finland and the international standards on thermal behaviour such as ISO 52016-1, EN 15316-2, and the EN 15251. The interlink between the power module that covers all appliances and the thermal model is presented in Figure 3.

While the building physics are taken from the international standards, some added values are supplemented to address the issue of the internal gains, including human metabolism and appliance usage, as well as interconnecting the ventilation with the thermal losses and running of the heating system. The occupancy scenario, as defined in section 2.1.2, defines further the level of activities and allows recalculating the metabolism from the inhabitants. In

---

[2] http://www.tracebase.org
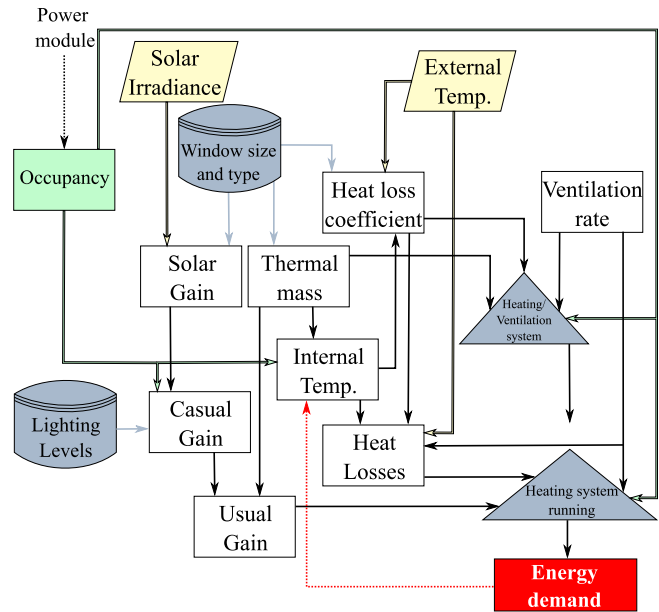
Figure 3. SBuM architecture connecting the power demand model and the heat demand model.
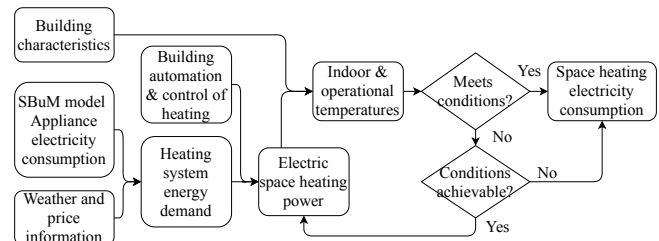
Figure 4. The operation of the electrical space heating part of the thermal model (Pulkkinen and Louis, 2019).

case of different type of space heating, the energy demand for each time resolution is provided and may potentially be used to recalculate the indoor temperature and electric heating power according to the predefined automation and control strategy. This operation principle is described in a flow sheet on the operation of the electric space heating part of the model presented in Figure 4 and in Pulkkinen and Louis (2019).

On-going work is being set up to include the use of domestic hot water provision from electric water tank .

2.2.3 Optimisations

The SBuM has a sub-block dedicated to the control system of the building and can consider multiple data input to optimise the consumption of the building. So far, two objectives are set, which are related to costs and environmental footprint of the consumed electricity. Secondly, the optimiser is only used when a home energy management system is setup for the studied house with the degree of control given to it. This means that the model can primarily be used to generate load profiles at a relatively high details, and can be consequently used in assessing optimisation control strategy based on dynamic pricing or other

power network indicators. As the SBuM model is based on Markov-Chain, it is inherent to model consequential actions and therefore disrupt the power cycle from appliances or heating units.

Objective functions:

$$\min \ p_{total,t} = \left[\left(\sum_{i=1}^{n} P_i \right) + P_{heat} + P_{ventil} \right] \times p_t \qquad (1)$$

where p is the energy expenditure from the house [€], and P is the power consumption from all appliances i, heating and ventilation [kWh], and for each price at time t [€/kWh].

2.3 Output data

Four main variables are obtained from the model: the total power consumption from the building that can be disaggregated by appliances and heating modes, thermal performances of the building including indoor temperature variation, heat demand, thermal comfort and their associated variables, the environmental impact, granted that the electricity mix from the grid is known as a dynamic real-time environmental impact, and the cost of these implementations.

3 Down-scaling simulation

Handling multiple time resolution from input to output of the simulation is a challenge. The time resolution must be adapted for flexible input data, flexible time optimisation, and flexible time resolution of the output signal from the model.

For this reason, the model was made flexible to have any sort of inputs with a maximum time resolution of one hour and some pre-defined time resolution for the input database: 30-15-3-1 minutes and 10 seconds. These resolutions were taken arbitrarily to match the existing databases of appliance signatures (see section 2.2.1), weather data resolution, and power status from the transmission network (down to 3 minutes per sample).

3.1 Scaling and multiplying

Down-scaling power and heat demand profiles, can be performed either as disaggregated profile from the modelled profile or built up from lower timestep resolution and appliance signatures. Both ways provide output with a high granularity and can be used further in power system modelling. The advantage of proceeding power disaggregation is to perform faster simulation with a larger timestep e.g. 1 hour. The downside is that control cannot be performed at a lower resolution than the model itself and therefore demand response and control will be performed using the simulation steps and not the disaggregated granularity. The other way of implementing a high resolution model is to increase the simulation step of the simulation and thus provide better control for demand response actions based on price or power network status. The SBuM

model allows carrying out both options with any simulation time. This is done by re-sampling the input dataset to meet the simulation time resolution and could be done simply by interpolating values, like in the case of temperature that does not vary very fast, or by performing linear interpolation and integrate a normally distributed noise to the dataset.

If a device is used for a very short time ($t_{cycl,n} < t_{min}$), then the studied appliances n could multiply its usage in order to comply with the average weekly usage of this device. This action multiplier only applies to new actions that has been triggered during the current iteration. Only in this condition the multiplying factor t_{mult} can be generated, otherwise its value is set to 1.

$$t_{cycl} = t_{mult} \times t_{cycl}$$
$$\text{where, } t_{mult} = \left\lfloor R \sim U \left[1, \frac{t_{min}}{t_{cycl}} \right] \right\rfloor \qquad (2)$$

T_{multi} is a dimensionless value multiplying the number of actions within a time step when needed to reach the weekly average usage of appliances. The value of T_{multi} is randomly generated and its value is comprised between a fraction of one cycle in comparison with the time step. The spread of the T_{multi} can be observed in Figure 5.
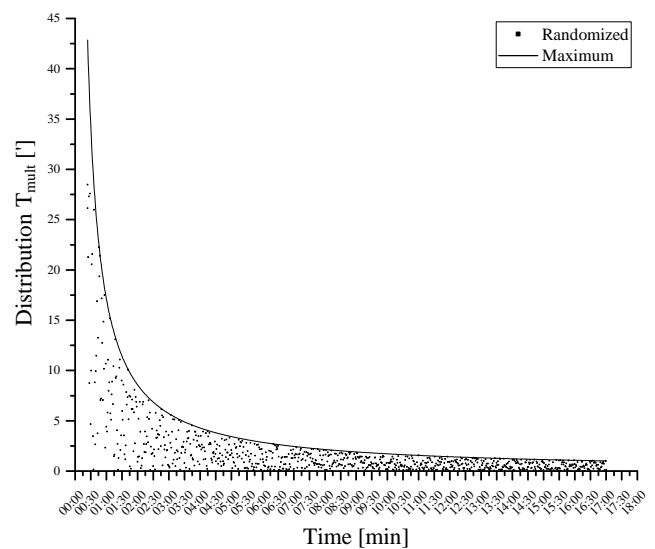
Figure 5. Distribution of t_{mult} at multiple variable time usage t_{cycl} in hourly timestep simulation.

4 Results and discussions

The results of the SBuM model will be reviewed in terms of model performance in generating load and production profiles. These profiles are generated at different time scales through an event distribution throughout a typical day (Figure 6). This allows comparing the input distribution profile per appliance with the generated distribution profile for the same appliances. Generated load profiles are also looked through the prism of a detailed profile and comparing generated dataset at different granularity (Figure 7). Results are compared with models proposed to the
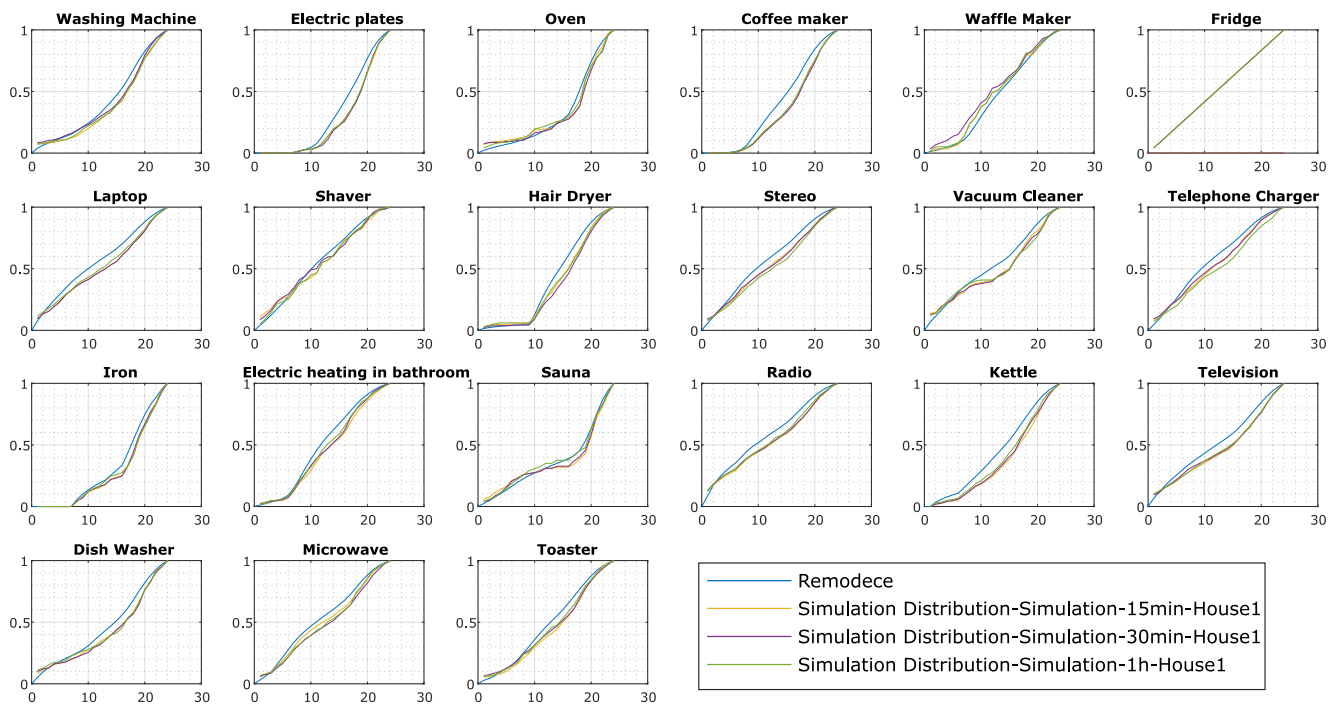
Figure 6. Cumulative daily distribution throughout the year against the REMODECE distribution used as input distribution in the SBuM model.

general public but do not allow high resolution modelling nor control options. For analysing the results, a set of 1-year simulation was performed for a test reference year, from January 1st to December 31st, 2012.

4.1 Power demand

The power demand of the buildings generated with the SBuM model will first be looked into the robustness to produce results similar to the input profiles given to the building, and second to the comparability levels of the different time resolution dataset produced while measuring their homogeneity through e.g. their daily load profiles.

4.1.1 Power profile distribution

Profile distributions for each appliance are taken from open dataset as described in section 2.2.1. The generated profiles are therefore easily comparable as their input profile are taken from the literature. A summary of 21 appliances used in the simulation is retrieved in Figure 6. The distribution profiles are aggregated to a daily load profile considering equally every day of the simulation (366 days).

Multiple statistical test were performed and for all appliances, the F-test are not statistically different at the 0.05 level, meaning that all distribution profile are statistically similar to the distribution dataset used for the appliances. This statistics is valid for all simulation granularity. In other words, the fit of distribution between the simulation output power profiles and the measured dataset of appliances measured in real living environment, demonstrate the robustness of the model and do not create unrealistic

scenarios where appliances are run randomly.

4.1.2 Dataset comparison

Power demands from the appliances are depicted for different time scale simulations in Figure 7. The power profiles are put in perspective with the active occupancy scenario that are drawn from the power consumption profiles. As the simulations are independently run and rely on normally distributed random numbers under daily distribution constraints, the dataset with different granularity does not run the same scenario. Nevertheless, they do present similarities in their daily routines. The power disaggregatation of the 15 minutes time resolution into a 10 seconds profile is also shown on Figure 7.

Although the distribution of the appliances is done with an hourly distribution, smaller time resolution allows disrupting the hourly distribution profile and insert actions within these time-frames. The good correlation between the active occupation and the inactive occupation e.g. sitting, lying or else, is due to the fact that occupation is deduced from the appliance usage. This is critical for the good interrelation with the thermal model that calculates the thermal comfort and the exchange of heat within the building from human occupation.

The total power consumption for the different runs are consistent where the 1h time resolution provides a power demand from appliances and lighting of 5050 kWh/y, the 30 minutes granularity accounts for 5293 kWh/y, and the 15 minutes resolution for 5158 kWh/y. In this particular case, the consistency is of the annual power demand between time scales were a critical indicators. The dataset
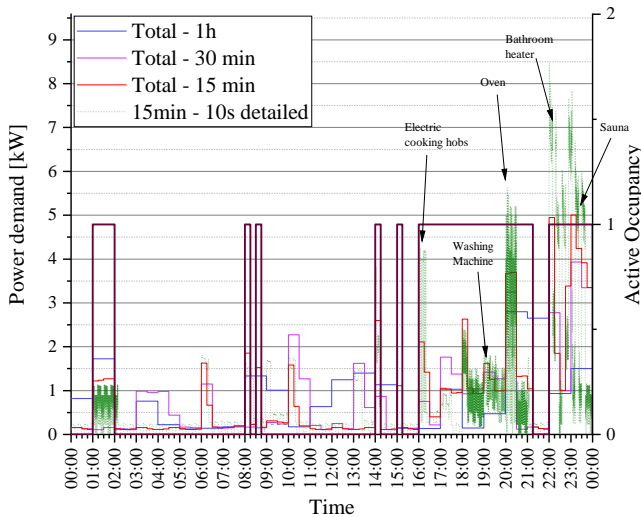
Figure 7. Energy service power load profiles and its occupancy for a typical summer day (1 day).

presents a similar mean value of power consumption for all time resolution simulations but with the higher granularity the power demand becomes more volatile which can be observed in the increase in their standard deviations and their associated coefficients of variation (1.19, 1.25, 1.30 for the 60 minutes, 30 minutes, and 15 minutes resolution respectively).

The daily load profile, which is intertwinked between the appliances power demand and their aggregation is depicted in Figure 8. The daily load profiles across the time
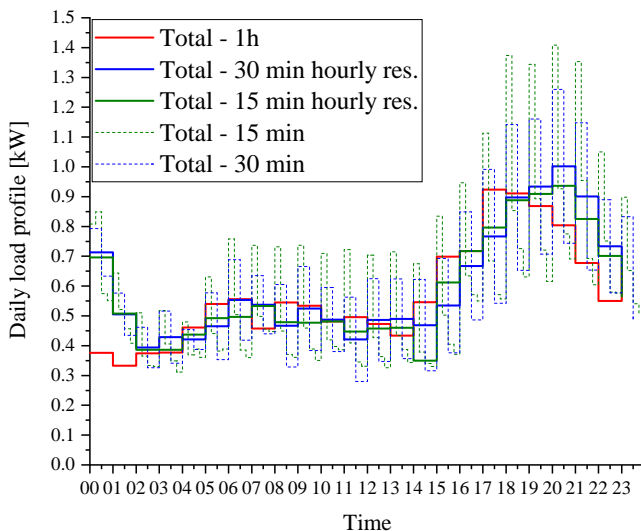
Figure 8. Mean daily load profiles generated by the 60/30/15 minutes time resolution simulations - one year simulation.

scales show good similarities that express their uniformity through the different runs of the model where the means for the 60 min, 30 min, 15 min are equal to .57 (SD = .17), .60 (SD = .18), and .59 (SD = .18) kW respectively. Figure 8 also depicts the daily load profile of the same building that has a higher simulation time resolution, presenting

therefore the higher variation. These variations are mainly due to the cycling of white appliances like the fridge that has a regular cycling time.

Downscaling the model time step to 10s allows fast control of some appliances if necessary. However, most of the benefits from existing demand side management (DSM) that reacts mainly on real time pricing that has been available at an hourly level so far. As there is intentions to decrease the resolution of day-ahead pricing to 15 minutes, it thus makes sense to be able to simulate buildings with this resolution and include their related controls. In any case, time resolution is always available at a 10 seconds resolution for the power module while the model may run with a larger time step, which can be meaningful for all sorts of engineering work. Furthermore, downscaling must be justified as it is a time consuming process where the computational time increases exponentially with the following relationship:

$$y(x) = \frac{1}{7.04} \cdot e^{\left(\frac{0.9+x}{7.04}\right)} \tag{3}$$

where x is the downscaling ratio. A full year simulation at an hourly time step takes about 6 minutes on a 32Gb RAM and i7 core, while the same simulation with a time step of 1 minute may take over 80 hours with the same hardware.

4.2 Thermal demand

Thermal load is also detailed at the chosen time resolution and shows more discrepancies between the model resolution while staying within the constraints set in the simulation. The thermal performance of the building is analysed through the total heat demand in a year, the variation of the operative temperature (weighted average of the indoor and surface temperatures as defined in the ISO 52016-1), and the power demand in terms of power variation per degree change in indoor temperature before heating the building. The data are represented in Figure 9 for the coldest week performed in the simulation (from February 3rd to February 10th). The temperature setting set in the simulation
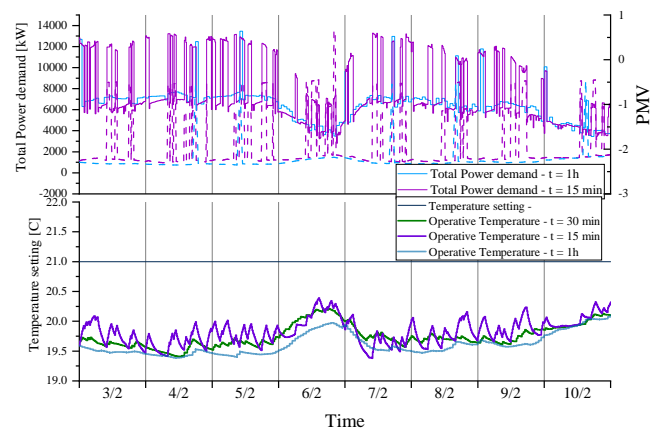
Figure 9. Indoor temperature variation and total thermal demand excluding DHW for the coldest week in the modelled year for three resolutions. Dotted lines represent the PMV variation for this same week

is 21°C and as it can be seen in the Figure 9, the operative temperature is varying below this constraint. The mean operative temperature of the building is very similar in all 3 simulations, with a mean operative temperature of 20.4°C (SD = .48), 20.5°C (SD = .52), and 20.6°C (SD = .7) for the 60/30/15 minutes simulations respectively. The heat demand for the 60/30/15 minutes simulation are 21.3 MWh/y, 23.4 MWh/y, and 24.4 MWh/y respectively. One of the reasons for difference of thermal energy consumption between the time resolutions is the shorter response time for adjusting the temperature within the building and keeping it to the correct level (adjusting the operative temperature to meet the temperature set for the house and therefore guaranteeing an indoor temperature of 21°C).

The thermal comfort varies more in the negative values with an average of -2 (SD = .28), -1.86 (SD = .45), and -1.82 (SD = .5) for the 60/30/15 minutes simulations respectively. Although the results indicate a cold feeling from the tenants (<-1), the main reason resides in a paradoxical setting for the heating system that is controlled to keep the indoor temperature at 21°C, and keeping the Predicted Mean Vote (PMV) level above (>) -.7 in the worst case. To reach the correct PMV level, a temperature setting of 23°C would be required but that would go against the Finnish regulation in its building code.

4.3 PV generation

The solar PV panel module presents results also for the different timestep sets for each simulation. The PV-module is compared against results from PVGIS EU-service that provides hourly resolution for different types of PV systems. The PV model module was tested with the global irradiation array used in PVGIS with a system loss factor of 14 %, which includes dirt and performance decrease over time, which is fixed through the simulation. The nominal power of the PV-module is 10 kW in both settings. Both results are presented in Figure 10 top left graph for a typical summer week. While the datasets are statistically not different at p = .01 during this extracted period and thus shows the good behaviour of the model during the summer months, the overall power generation from PV panels are different at p = .01 for the overall year. In other words, the behaviour of the PV model is validated with some small discrepencies in the winter months resulting in different overall yearly energy production. The overall yearly power produced is however different with a total of 12.9 MWh/year for the PVGIS model and 14.1 MWh/year for the SBuM model. The main difference resides in the winter month when the production is low, especially under high latitude where the simulations were performed (65.05° North).

When running the simulation with more precise dataset from the Finnish Meteorological Institute (FMI), the power production has a greater Coefficient of Variation (CV = 1.52 for PVGIS and CV = 1.84 - 1.9 with FMI dataset) and this is due to the higher variation in the global irradiance dataset provided by FMI. The total production
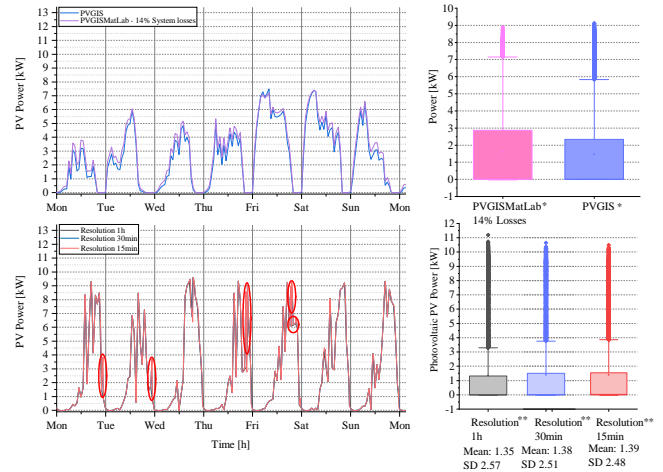
Figure 10. *Top:* Week production from a 10 kW PV system run with PVGIS online tool and the SBuM model using the PVGIS solar radiation data. *Bottom:* The SBuM model output using the Finnish Meteorological Institute solar radiation data for 3 time resolution, 1h, 30 minutes, and 15 minutes. *The population mean is significantly different at p = 0.01, **The population means are not significantly different at p = 0.01.

from the PV-panels using the FMI dataset is 11.9 MWh/y, 12.1 MWh/y, and 12.2 MWh/y for the 60/30/15 minutes time resolution respectively. The model output is thus consistent with state of the art global output variables while providing more detailed information while the simulation resolution is increasing without altering the global output of the model, thus guaranteeing its use for automation purpose.

5 Conclusions

This paper presented a multi-scale open source modelling tool to evaluate the impact of smart control on the power and heat demand of the building. The SBuM model includes both the heat and power demand, as well as thermal comfort user satisfaction and willingness to shift load from appliances in case of energy management system. The model can be run at different time resolution from 60 minutes down to 10 seconds, therefore, providing a wide range of possible control strategies for distribution system operators as well as provide information to homeowners. As the SBuM model uses a Markov-Chain approach to the modelling of the buildings, the different energy points can be managed on the fly but post processing of the results can be done for further use. The real challenge was to create synthetic profiles that are comparable with different time scale resolutions and flexible in terms of computing, to be included in market operation as well as automation solutions.

The SBuM model showed robustness in the creation of load profiles for both thermal and power demands. The SBuM model generated profiles and datasets that were comparable and statistically not different across multiple time resolutions. However, the smaller the time resolution, the greater is the computing time of the simulations,

therefore having a very low granularity in the simulation as it may take a very long time to compute when the simulation time horizon is set to a year for instance.

In the future, the SBuM model will integrate a model for domestic hot water integration with controllability in case it is electrically heated. It will therefore cover the whole range of energy consumption in dwellings. Further, the impact of different market configuration for dynamic pricing will be investigated to evaluate the impact of and on buildings. Heating control strategies will need to be implemented further to pilot electric convectors in a way that it does not increase the total energy consumption of the house.

Acknowledgement

References

Aníbal De Almeida and Paula Fonseca. Residential monitoring to decrease energy use and carbon emissions in Europe. *International Energy . . .* , pages 1–14, 2006. URL `http://www.isr.uc.pt/{~}remodece/news/Paper{_}DeAlmeida.pdf`.

Drury B. Crawley, Linda K. Lawrie, Frederick C. Winkelmann, W.F. Buhl, Y.Joe Huang, Curtis O. Pedersen, Richard K. Strand, Richard J. Liesen, Daniel E. Fisher, Michael J. Witte, and Jason Glazer. EnergyPlus: creating a new-generation building energy simulation program. *Energy and Buildings*, 33(4):319–331, Apr 2001. doi:10.1016/s0378-7788(00)00114-6.

EU Commission. The european green deal, December 2019. URL `https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52019DC0640`.

Søren Østergaard Jensen, Anna Marszal-Pomianowska, Roberto Lollini, Wilmer Pasut, Armin Knotzer, Peter Engelmann, Anne Stafford, and Glenn Reynders. IEA EBC annex 67 energy flexible buildings. *Energy and Buildings*, 155:25–34, Nov 2017. doi:10.1016/j.enbuild.2017.08.044.

Jean-Nicolas Louis, Antonio Caló, Kauko Leiviskä, and Eva Pongrácz. Modelling home electricity management for sustainability: The impact of response levels, technological deployment & occupancy. *Energy and Buildings*, 119:218–232, May 2016. doi:10.1016/j.enbuild.2016.03.012.

Antonio Luque and Steven Hegedus, editors. *Handbook of Photovoltaic Science and Engineering*. John Wiley & Sons, Ltd, Dec 2010. doi:10.1002/9780470974704.

Eoghan McKenna and Murray Thomson. High-resolution stochastic integrated thermal–electrical domestic demand model. *Applied Energy*, 165:445–461, Mar 2016. doi:10.1016/j.apenergy.2015.12.089.

David Murray, Lina Stankovic, and Vladimir Stankovic. An electrical load measurements dataset of united kingdom households from a two-year longitudinal study. *Scientific Data*, 4(1), Jan 2017. doi:10.1038/sdata.2016.122.

D.M. Murray, J. Liao, L. Stankovic, and V. Stankovic. Understanding usage patterns of electric kettle and energy saving potential. *Applied Energy*, 171:231–242, Jun 2016. doi:10.1016/j.apenergy.2016.03.038.

Fabio Orecchini and Adriano Santiangeli. Beyond smart grids – the need of intelligent energy networks for a higher global efficiency through energy vectors integration. *International Journal of Hydrogen Energy*, 36(13):8126–8133, Jul 2011. doi:10.1016/j.ijhydene.2011.01.160.

Fabiano Pallonetto, Mattia De Rosa, Francesco D'Ettorre, and Donal P. Finn. On the assessment and control optimisation of demand response programs in residential buildings. *Renewable and Sustainable Energy Reviews*, 127:109861, Jul 2020. doi:10.1016/j.rser.2020.109861.

Jari Pulkkinen and Jean-Nicolas Louis. Utilizing flexibility of electric heating in demand side management programs in Finland in 2050. In *Proceedings of 11th International Conference on Applied Energy*, volume 4 (3), 2019.

Andreas Reinhardt, Paul Baumann, Daniel Burgstahler, Matthias Hollick, Hristo Chonov, Marc Werner, and Ralf Steinmetz. On the accuracy of appliance identification based on distributed load metering data. In *Proceedings of the 2nd IFIP Conference on Sustainable Internet and ICT for Sustainability (SustainIT)*, 2012.

Melody Stokes. *Removing barriers to embedded generation : a fine-grained load model to support low voltage network performance analysis*. PhD thesis, Institute of Energy and Sustainable Development de Montfort University, Leicester, 2005. URL `http://hdl.handle.net/2086/4134`.

Andong Wang, Rongling Li, and Shi You. Development of a data driven approach to explore the energy flexibility potential of building clusters. *Applied Energy*, 232:89–100, Dec 2018. doi:10.1016/j.apenergy.2018.09.187.

Selin Yilmaz, Steven K. Firth, and David Allinson. Occupant behaviour modelling in domestic buildings: the case of household electrical appliances. *Journal of Building Performance Simulation*, 10(5-6):582–600, Feb 2017. doi:10.1080/19401493.2017.1287775.

Modelling and Evaluation of an Adiabatic Compressed Air Energy Storage System with Packed Beds

Marianne Fjellestad Heitmann   Carlos Pfeiffer   Elisabet Syverud

Department of Process, Energy and Environmental Technology, University of South-Eastern Norway, Norway,
marianneheitmann94@gmail.com , Carlos.pfeiffer@usn.no , Elisabet.syverud@usn.no

Abstract

Compressed Air Energy Storage (CAES) is a promising alternative for energy storage. An Adiabatic Compressed Air Energy Storage (A-CAES) system has been analysed in this paper, to store excess energy production from a wind turbine generator for up to one week. Compressed air is stored in a cavern of constant volume. The heat produced by the compression of the air during the charge process is stored in several packed beds and used later during discharge to reheat the air prior to energy production.

This paper presents a preliminary thermodynamic analysis estimating the size of the system for a given quantity of energy storage, a dynamic model including packed beds for additional energy storage, and a simulation made in MATLAB to analyse the efficiency of the system. The A-CAES roundtrip efficiency is 53%. The physical limitations of the actual compressors and expanders were taken into account when defining the operational pressures and temperatures in the model. A preliminary capital cost estimation of the system was conducted for Norway, resulting in an estimated investing cost of approximately 2 700 NOK/kWh.

Keywords: compressed air energy storage, adiabatic, thermal energy storage, packed bed, thermodynamic analysis

1 Introduction

The use of renewable energy resources has increased in the past years. In Norway, the energy source that increases the most is wind energy (Øvrebø, 2020). This is an intermittent energy source, so efficient energy storage is needed (Grazzini and Milazzo, 2008). Compressed Air Energy Storage (CAES) Systems are promising for storing energy. In a CAES the excess electrical energy is used to compress air to a high pressure. Then the air is stored in caverns or tanks. When the energy is needed, the pressurized air is released, reheated, and expanded in turbines. The turbines drive a generator that produces electric energy. This energy is then returned to the grid (Eckroad, 2003).

CAES systems typically have a storage capacity of 5MWh – 1GWh, with a discharge time from a couple of hours to days (Moore and Shabani, 2016). In a Diabatic CAES (D-CAES), the air needs to be reheated to produce electric energy. The reheating is done by burning fossil fuels (Calero et al., 2019). This is not a desired solution, because these systems are not CO_2 neutral, and the fuel brings an increase in the operational cost. The Adiabatic CAES (A-CAES) system does not need additional fuel consumption and is the system studied in this work. In this paper, we provide a preliminary thermodynamic analysis estimating the size and cost of an A-CAES for a given quantity of energy storage and using the physical limitations of actual compressors and expanders.

1.1 Adiabatic CAES

In an Adiabatic Compressed Air Energy Storage system (A-CAES), thermal energy storage (TES) replace the combustor of the D-CAES. Energy is expelled as heat during the compression stage. This heat is stored in TES and later used to reheat the air during the expansion stage. This contributes to both higher efficiency of the plant, as well as zero CO_2 emissions from fuel. There has been a lot of research on this type of system but so far only to the stage of simulation of models and theoretical analysis (Yang et al., 2014).

ADELE is a German project where the goal is to build an A-CAES system with a roundtrip efficiency of 70%. When the air is compressed, the heat is captured in a heat-storage facility instead of being released to the surroundings. Later during the discharge, the energy in the heat-storage is released into the compressed air. With this solution, no fuel combustion is needed to heat the air. By avoiding the use of fossil fuel, this technology provides a CO_2 neutral delivery of the peak-load electricity from renewable energy (RWE Power, 2010).

The storage temperature of the plant is one of the most important parameters in an A-CAES. There are three main temperature categories: High temperature A-CAES operates above 400°C, Medium temperature A-CAES runs between 200°C to 400°C, while Low temperature A-CAES operates below 200°C (Degl'Innocenti, 2018). A-CAES simulations using a packed bed as the thermal energy storage component, report an efficiency of 70%, for a temperature range of 30°C to 430°C (Barbour et al., 2015). Wolf and Budt (2014) report a roundtrip efficiency of 52% to 60% for low-temperature A-CAES when the system operates at

temperature from 90°C to 200°C. Sciacovelli et al. (2017) found the roundtrip efficiency to be in the range from 60% to 70% when the TES system has an efficiency of 90% and the system operates at temperatures ranging from 100°C to 500°C.

2 System Overview

The A-CAES presented in this work is a three-stage system that includes compressors, packed beds, turbines, valves, and a cavern, Figure 1.

The cavern is charged by compressed air from the last stage of three compressors. The compressors are powered by surplus energy, e.g. from wind turbines. The hot air out of each compressor enters at the top of the packed bed and exits at the bottom. The packed beds have ceramic Raschig rings that cool down the air and store the thermal energy. The Raschig rings provide a large heat exchange surface allowing quick exchange of heat from the air to the packed bed. The temperature differences across the rings are very low (Heitmann, 2020).

The discharge process is similar, but in reverse order: when the air is released from the cavern, it re-enters the last packed bed, now to reheat the air. In this process, the air enters at the bottom with a lower temperature and exits at the top with a higher temperature due to release of the stored heat from the charge process. The heated air is expanded by a turbine, which drives an AC generator that generates electric energy. The expansion process cools the air, so the air enters the next packed bed to be re-heated. As in the charge process, this happens in three stages, but now the air is expanded in three turbines (one at the discharge of each packed bed). The electric energy generated by the turbines is released back to the grid.

The minimum operational pressure in the cavern is 40 bar, while the maximum pressure is 70 bar. When the pressure is below 70 bar, and there is surplus electric energy, the charge mode is activated. When the pressure in the cavern reaches 70 bar, the compressors stop. The discharge mode starts when the energy is needed. This will go on until there is surplus energy available, or the pressure in the cavern decreases to 40 bar.

Valves are used to control the flow of air, as shown in Figure 1. The check valve makes sure the flow of air only goes in one direction, while the control valves by the cavern regulate the air pressure during the charge and discharge modes. The valve at entry to the first expander keeps the pressure into the expander constant at 40 bar. For the air to flow to the cavern, each increment of air must be compressed to a pressure just above the storage pressure. The valve at the cavern entry ensures that the pressure upstream is always higher than the cavern pressure during the charge mode. A more detailed description of the system can be found in (Heitmann. 2020).

3 Dynamic Model

To analyse the system behaviour we developed a dynamic model based on Heitmann (2020, p. 20). The
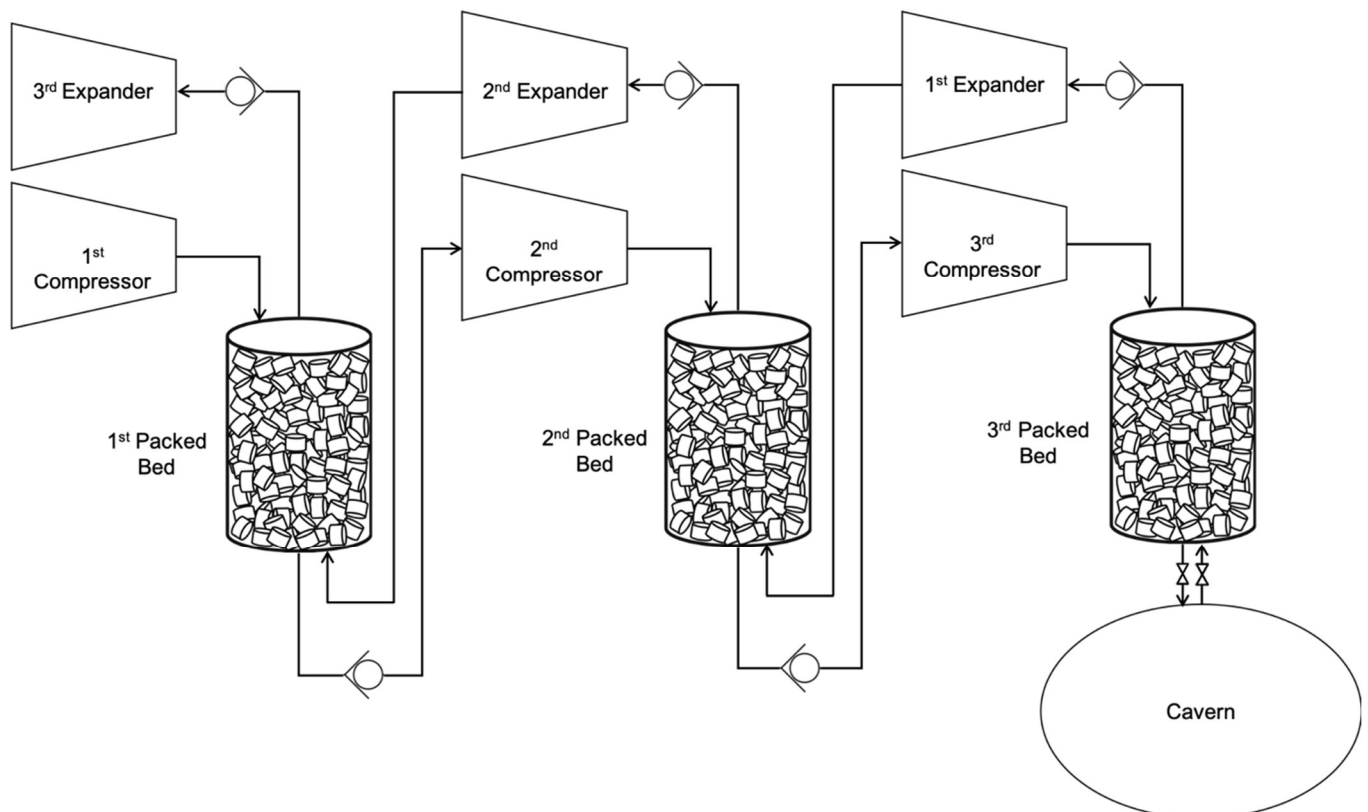
Figure 1. Sketch of the system simulated in this paper

model is described in this section.

3.1 Adiabatic Compression

For an ideal, reversible adiabatic compressor, and assuming ideal gas, the final temperature is given by:

$$T_1 = T_{in}\left(\frac{p_1}{p_{in}}\right)^{\frac{R}{c_p}} \qquad (1)$$

where c_p is the specific heat capacity for constant pressure (kJ/kmol K), R is the universal gas constant (8.314 kJ/kmol K), T_1 is the final temperature (K), T_{in} is the temperature of the air into the compressor (K), p_1 is the pressure after compression (kPa), p_{in} is the pressure before the compression (kPa).

The actual work used by the compressor is calculated by Eq.(2):

$$W_{C1} = \dot{m}\, c_p(T_1 - T_{in})\frac{1}{\eta_{isen}} \qquad (2)$$

where η_{isen} is the isentropic efficiency of the compressor.

3.2 Packed Bed

The conservation of energy is used to express the equations for the temperature of the air and the solid phase in the packed bed:

$$\Delta E = \dot{m}_{in}\Delta\hat{H} - \dot{m}_{out}\Delta\hat{H} - \hat{h}\, A\, (T_f - T_s) \qquad (3)$$

where $\Delta\hat{H}$ is the change in enthalpy (kJ/kmol).
On the right-hand side, the first two terms are the net heat input due to the fluid flow and the last term is the heat transfer to the solid.

The packed bed is used to cool down the air so the temperature of the air will be higher at the top of the bed than the bottom. To make the calculations of the temperature distribution simpler, the bed is divided into n sections. When assuming constant mass flow $\dot{m}_{in} = \dot{m}_{out} = \dot{m}$, the energy rate for the air in a section is:

$$\varepsilon A\rho_a c_{p,a}\frac{dT_a}{dt}\,\Delta z = -v_a\rho_a A c_{p,a}\frac{dT_a}{dz}\Delta z - \hat{h}_{vol}A\Delta z\,(T_a - T_s) \qquad (4)$$

where, ε is the void fraction of the air, A is the area (m²), ρ is the density (kg/m³), $\frac{dT_a}{dt}$ is the change of the temperature of the fluid over time, Δz is the height of the section (m), v is the velocity (m/s), $\frac{dT_a}{dz}$ is the change of temperature of the fluid over the length (K/m), $\hat{h}_{vol}$ is the volumetric heat transfer coefficient (kJ/m³ s K), and T is the temperature (K). Subscript a stand for air and s for solid. For the solid the energy rate is:

$$(1 - \varepsilon)A\Delta z\rho_s c_{p,s}\frac{dT_s}{dt} = -\hat{h}A\Delta z\big(T_f - T_s\big) - \varepsilon\, A\frac{d}{dz}\left(\lambda_s\frac{dT_s}{dz}\right) \qquad (5)$$

The last term of Eq. (5) is the heat conduction through the solids in the packed bed. The contact between the solids are very small, so this term can be neglected. The equation becomes:

$$(1 - \varepsilon)\rho_s c_{p,s}\frac{dT_s}{dt} = -\hat{h}_{vol}\big(T_f - T_s\big) \qquad (6)$$

To calculate the volumetric heat transfer coefficient, $\hat{h}_{vol}$, between the packing elements (for this work Raschig rings were considered) and the air, the empirical relationship suggested by (Coutier and Farber, 1982) was used:

$$\hat{h}_{vol} = 700\left(\frac{G}{d_p}\right)^{0.76} \qquad (7)$$

where G is the core mass velocity of the air (kg/m²s), and d_p are the average particle diameter (m).

3.3 Flow of Air

The flow of air is calculated by: (Heitmann, 2020, p. 22)

$$Q_G = 962\, C_{V\,valve}\sqrt{\frac{p_1^2 - p_2^2}{SG\, T}} \qquad (8)$$

where Q_G is the gas flow in standard cubic feet per hour, $C_{V,valve}$ is the coefficient of flow, p_1 is the inlet pressure in psia, p_2 is the outlet pressure in psia, SG is the specific gravity of the air at 70°F and 14.7 psia, and T Is the absolute temperature in °R. None of these are SI units, so conversion equations are used to get the answer in the correct unit. These kinds of correlations are common to calculate air flow through a control valve, where the value $C_{V,valve}$ can be adjusted by modifying the aperture of the valve to get the required flow.

3.4 Cavern

Since the volume of the cavern is constant, the mass in the cavern will vary due to the pressure increase and decrease over time. The mass balance for the cavern are:

$$\Delta M = \dot{m}_{in}\Delta t \qquad (9)$$

where ΔM is the change in total mass (kg), $\dot{m}_{in}$ is the mass flow of air into the cavern (kg/s), and Δt is the change in time (s). Taking the limit when Δt tends to zero leads to:

$$\frac{dM}{dt} = \dot{m}_{in} \qquad (10)$$

The temperature of the cavern will vary as well as the total mass and pressure. The energy balance for the cavern for a small time Δt, neglecting changes in kinetic and potential energy of the air, is given by:

$$\Delta\left(M\,\widehat{U}\right) = \dot{m}_{in}(t)\widehat{H}_{in}(t)\Delta t \qquad (11)$$

Which becomes when Δt tends to zero:

$$\frac{d\left(M\,\widehat{U}\right)}{dt} = \dot{m}_{in}\widehat{H}_{in}(t) \qquad (12)$$

Where:

$$\frac{d\left(M\,\widehat{U}\right)}{dt} = M\frac{d\widehat{U}}{dt} + \widehat{U}\frac{dM}{dt} \qquad (13)$$

$\widehat{U}$ is the internal energy of the air already inside the cavern, and $\widehat{H}_{in}$ is the enthalpy of the air flowing into the cavern. $\widehat{U}$ can be calculated by

$$\widehat{U} = \widehat{U}_0 + c_v(T - T_0) \qquad (14)$$

where c_v is the specific heat capacity for constant volume (kJ/kmol K), $\widehat{U}_0$ is a reference internal energy (kJ/kmol) and T_0 is the reference temperature (K). The enthalpy can be calculated by:

$$\widehat{H}_{in} = \widehat{H}_0 + c_p(T_{in} - T_o) \qquad (15)$$

where $\widehat{H}_0$ is a reference enthalpy (kJ/kmol) at T_0.

When inserting eq. (17), eq. (18) and eq. (19) into eq. (16), it becomes:

$$M\,C_v\frac{dT}{dt} + \left(\widehat{U}_o + c_v(T - T_0)\right)\dot{m}_{in} = \\ \dot{m}_{in}(\widehat{H}_0 + c_p(T_{in} - T_0)) \qquad (16)$$

Eq. (20) can be simplified to:

$$\frac{dT}{dt} = \frac{\dot{m}_{in}(c_p T_{in} - c_v T)}{M\,c_v} \qquad (17)$$

3.5 Adiabatic Expansion

We use the energy balance for the expander to calculate the temperature of the air after expansion. This is similar to the energy balance for the compressor but since the work is extracted from the system, the work is negative.

For the temperature after the expander, the following equation is used: (Heitmann, 2020, p. 23)

$$T_2 = T_{in}\left(\frac{p_2}{p_{in}}\right)^{\frac{R}{c_p}} \qquad (18)$$

The work generated by the expander can be calculated by:

$$W_E = \dot{m}\,c_p(T - T_{in})\,\eta_{isen} \qquad (19)$$

where η_{isen} is the isentropic efficiency. Since the expanders are adiabatic, the ideal process is isentropic (Saravanamuttoo et al., 2017).

3.6 Efficiency of the System

The efficiency of the system η is calculated by:

$$\eta = \frac{W_{discharge}}{W_{charge}} \qquad (20)$$

where $W_{discharge}$ is the total work produced by the expanders during discharge and W_{charg} is the total work done by the compressors in the charge part.

4 Cost Estimation

To check whether the A-CAES system is economically viable, a cost estimation is needed. A perfect estimation is difficult to accomplish, but a simple estimation for the system based on data from other systems, the given size and the material used is shown in this chapter to give an idea of the magnitude of the cost.

4.1 Compressors and Expanders

We used the SIEMENS DATUM calculator (SIEMENS, 2020) to estimate the cost of the compressors. The DATUM compressors are common in the oil and gas industry. The tool give an initial cost estimates sufficient for preliminary analysis. The tool calculates a three-stage compressor with three impellers and intercooling. The CAPEX of one compressor are assumed to be 500 000 \$. According to Buffa et al. (2013) the cost of a Mixed Flow Expander is somewhat the same as for the compressor.

4.2 The Packed Beds

The cost estimation for the packed beds is based on the volume of the column and the mass of the solid.

The height and radius of each column are 30m and 5m, respectively. This leads to a volume of 2 355 m³. The columns are assumed to be inside the mountain, so the cost of mountain blasting needs to be taking into account. In Norway, mountain blasting is expensive and runs at about 200-300 NOK per cubic meter (NOK/m³) (Byggestart, 2020).

The solid inside the column are Raschig Rings in ceramic with a void fraction of 74%. The ceramic is Alumina (Al2O3), with a density of 3 900 kg/m³. The

cost of the ceramic is 1.90 $/kg (19.4 NOK/kg) (Heitmann, 2020).

4.3 The Cavern

The cost of the cavern is the mountain blasting, which is the same as for the packed beds, i.e. 200-300 NOK per cubic meter. The volume of the cavern is calculated to be 41 300 m^2.

4.4 Cost of the System

The total cost of the system is based on the calculation is calculated to be 175 MNOK. With a total capacity of 66 400 kWh, the capital cost estimate is 2 700 NOK/kWh.

This cost estimation only includes the main parts of the system: three compressors, three packed beds, three expanders and a cavern. The cost is the capital cost of the system, not including cost of pipes, valves, or insulation. A substantial amount of money can be saved if an existing cavern is used, instead of blasting one.

5 Simulation and Efficiency Results

5.1 Total Mass in the Cavern

The variations of the total mass in the cavern during the charge and discharge processes are shown in Figure 2 and Figure 3. The red line is the increase of mass in the cavern during the charge process. It starts at an initial mass of 2.03 x10^6 kg, which is calculated by assuming ideal gas at 40 bar and the air cave temperature. After about 30 hours (~108 000 s) the cavern reaches the maximum operational pressure (70 bar) and the maximum mass. When this value is reached the charge process stops.

The blue line shows the decrease of mass in the cavern during the discharge process. It starts at the maximum mass of 3.09 x10^6 kg, and after about 47 hours (~169 000 s) the cavern reaches the minimum operational pressure of 40 bar. The x-axis shows the time in seconds from 0 to 172800 s (48 hours).

5.2 Temperatures over the Packed Beds

The temperature distributions over the packed beds in the charge process are shown in Figure 4. Five sections (n=5) are used in the simulations. Originally, when the system is "loaded" for the first time, all the sections are at the same temperature. The green line shows the temperature of the air in the first section of the packed bed, corresponding to the section where the air enters the packed bed. The purple line shows the temperature of the air in the second section of the bed. The red line shows the temperature of the air in the third section. The pink one shows the temperature of the air in the fourth section. Finally, the turquoise line shows the temperature of the air in the last section of the packed bed. This turquoise line represents the air temperature

entering the next compressor stage or the cavern. The black dotted lines represent the temperature of the solid material in the corresponding section, the solid temperature closely follow the air temperature of all sections. The simulation covers 48 hours of operation. The charge process ends after about 30 hours because the cavern is fully loaded with air at 70 bar. This means that the flow of air stops and the temperature of the air and the solid in the packed beds will stop increasing. We assume well-insulated packed beds where the temperatures stay constant when there is no air circulating through the system.
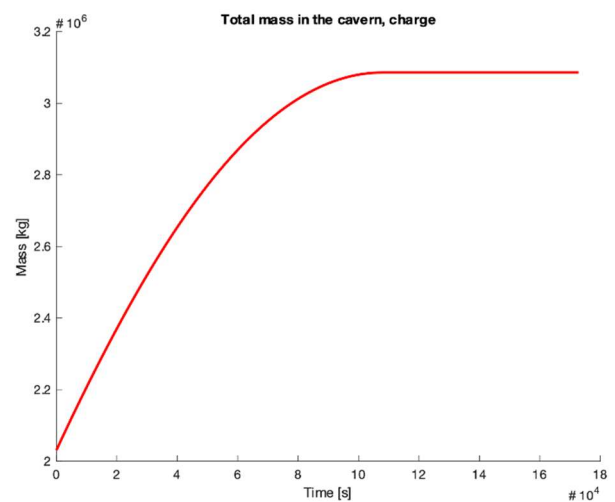
Figure 2. Total mass in the cavern during the charge process.
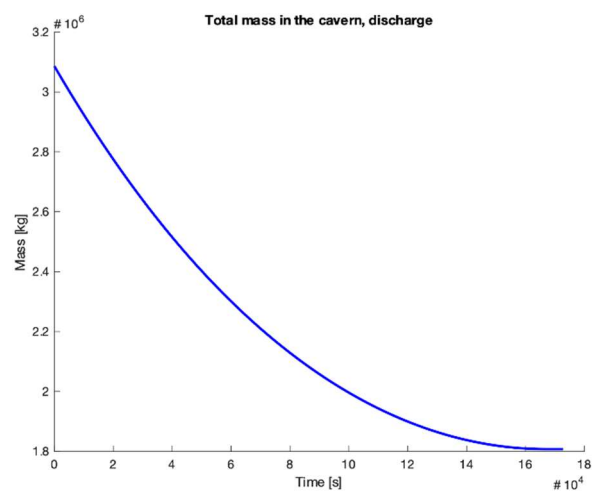
Figure 3. Total mass in the cavern during the discharge process.

a)

b)

c)
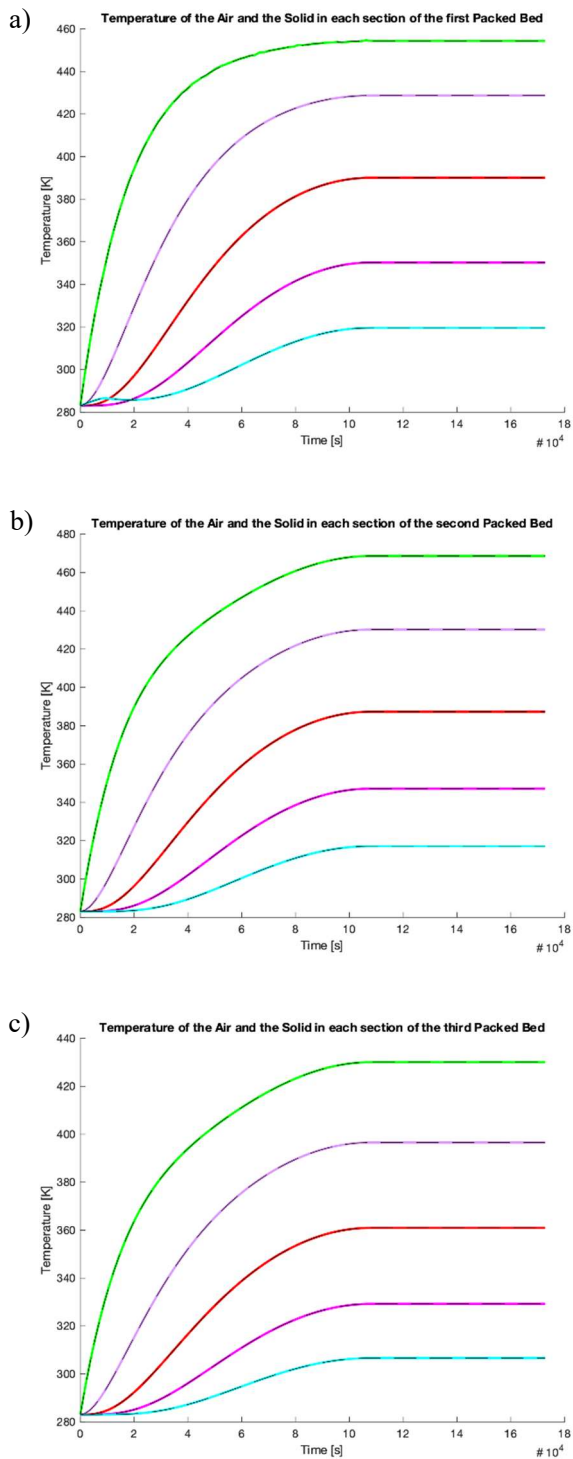
Figure 4. Calculated temperature over the packed beds in the charge process: a) is the first packed bed, b) is the second packed bed, and c) is the third packed bed.

a)

b)

c)
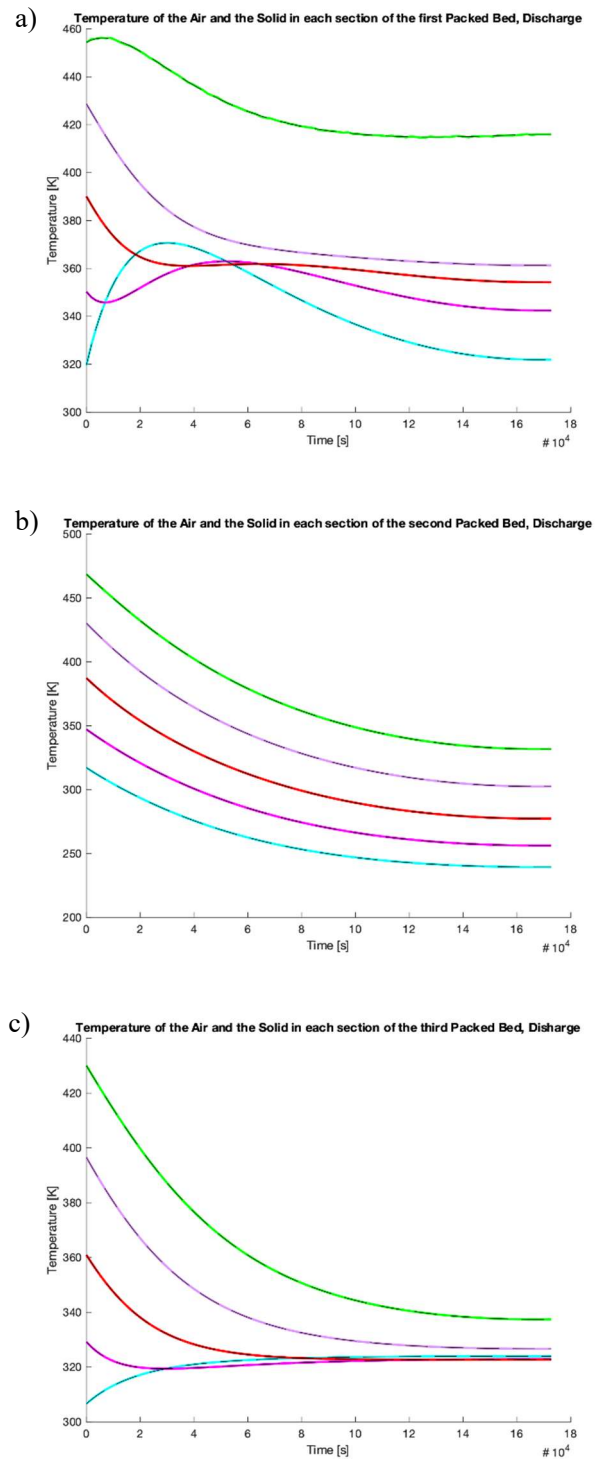
Figure 5. Calculated temperature over the packed beds in the discharge process: a) is the first packed bed, b) is the second packed bed, and c) is the third packed bed.

Figure 5 shows the temperatures over the packed bed during discharge. The initial values for all the sections are the end temperatures reached during the charge process. In the discharge, the first packed bed is the last one from the charge process. The air from the cavern enters at a temperature of 50°C at the bottom of the packed bed and is heated through the bed.

The turquoise line shows the temperature of the air in the bottom of the bed, which is the first section in the discharge process. The pink line shows the temperature of the air in the next section of the bed. The red line shows the temperature of the air in the middle section of the bed. The purple line shows the temperature of the air in the fourth section. Finally, the green line shows the temperature of the air at the top section of the packed bed. This is the air temperature leaving the packed bed and entering the expander. The black dotted lines represent the temperature of the solid in the corresponding section. The solid material temperature closely follows the air temperature in the packed bed.

5.3 System Efficiency

The calculated work done by the three compressors during the charge process was 2 590 kW.

The work produced by the expanders during discharge process was 1 380 kW. This results in an efficiency of 53.4 %. This result indicates that more than half of the electric energy used to compress the air is regenerated in the form of electric energy, without the use of additional fuel to reheat the air.

6 Discussion

The model for the charge process works well and provides realistic results. The air is cooled down to a temperature acceptable for the next compression. The packed bed material is heated fast and reaches the temperature of the compressed air.

The discharge requires operation parameters adjustments to get realistic practical temperatures for the discharge air. According to the simulation, the output temperature for the first expander is between 230 K and 290 K. We consider this temperature range as too low to provide an efficient output from a real turbo expander.

The size of the packed beds was the same in the simulation, but it could be interesting to see how the system would behave if different bed sizes are fitted to the compressors in each stage. Also, the simulation was only done for one type of packing material, ceramic Raschig rings. The system could be tested with other materials, such as metal Raschig rings or marbles to find the solution that provides the highest efficiency with the lowest cost.

In this paper, we have simulated only one cycle (one charge and one discharge). It would be interesting to see how the temperatures over the packed develop with multiple cycles and if the efficiency would be affected.

Although the physical limitations of the components were taken into account, the maximum compressor discharge temperatures are lower than the minimum inlet temperature allowed in most expanders or turbines. This can make it difficult to use existing equipment. For this system to work in reality, special equipment may be needed.

The reviewed literature for similar systems reports an efficiency between 60 - 70 %. Even though the results obtained in our simulations are lower (53.4%), they are similar to the system efficiency of real existing systems using additional fuel to reheat the air is on the range of 40% to 54% (Luo and Wang, 2013). Furthermore, our proposed A-CAES system does not have CO_2 emissions.

The system cost is calculated at 2 700 NOK/kWh. There are not a lot of numbers to compare with. However, in the A-CAES presented by Mignard (2014) the cost is estimated at 334 000 USD/MWh, (3 400 NOK/kWh) in a similar but smaller capacity system using steel as the packing material.

7 Conclusion

We used thermodynamic analysis, modelling, and simulation to make a preliminary estimate of the size of a 12 MW A-CASE system with a storage time of 48 hours. The size of the cavern was calculated to be 41 300 m^3. This value is calculated for a perfect reversible process, so this size is expected to be larger for an actual system.

A dynamic model was created for the system, including the packed beds, the mass and temperature of the cavern, and the work consumed and produced. The model was simulated in MATLAB. We calculated the work done by the compressors as 2 590 kW and the work produced by the expanders as 1 380 kW, which corresponds to a capacity of 66 400 kWh. The system roundtrip efficiency was calculated at 53.4%. The total cost is calculated to be 175 MNOK. With a total capacity of 66 400 kWh, the capital cost estimate is 2 700 NOK/kWh.

We find that the system discharge temperature is too low for use with existing expander technology. In addition, we find that the cost of A-CAES is too high for the system to be economically viable. The cost is site specific and depends on local conditions of the air reservoir and the cavern is by far the most expensive component.

References

E. Barbour, D. Mignard, Y.Ding, and Y. Li. Adiabatic Compressed Air Energy Storage with packed bed thermal energy storage. *Applied energy,* 155: 804-815, 2015. doi:10.1016/j.apenergy.2015.06.019.

F. Buffa, S. Kemble, G. Manfrida, and A. Milazzo. Exergy and Exergoeconomic Model of a Ground-Based CAES

Plant for Peak-Load Energy Production. *Energies, 6*(2):1050-1067, 2013. doi:10.3390/en6021050.

Byggestart. *Graving: Hva koster det? (2020-pris)*. Retrieved from www.byggstart.no/pris/graving. Accessed 8-May-2020.

I. Calero, C.A. Cañizares, and K. Bhattacharya. Compressed Air Energy Storage System Modeling for Power System Studies. *IEEE transactions on power systems*, 34(5): 3359-3371, 2019. doi:10.1109/TPWRS.2019.2901705.

J.P. Coutier and E.A. Farber. Two applications of a numerical approach of heat transfer process within rock beds. *Solar energy, 29*(6):451-462, 1982. doi:10.1016/0038-092X(82)90053-6.

J. Degl'Innocenti. *Compressed air energy storage for clean offshore energy supply*. Norges Teknisk-Naturvitenskapelige Universitet, Master's thesis, 2018.

S. Eckroad, *EPRI-DOE Handbook of Energy Storage for Transmission & Distribution Applications*, Report 1001834, 2003. Retrieved from www.sandia.gov/ess-ssl/publications/ESHB%201001834%20reduced%20size.pdf. Accessed 8-May-2020.

G. Grazzini and A, Milazzo. Thermodynamic analysis of CAES/TES systems for renewable energy plants. *Renewable energy*, 33(9):1998-2006, 2008. doi:10.1016/j.renene.2007.12.003.

M.F. Heitmann. *Modelling and Evaluation of Compressed Air Energy Storage System*. University of South-Eastern Norway, Master's thesis, 2020.

X. Luo and J.Wang. *Overview of Current Development on Compressed Air Energy Storage*. University of Warwick, Coventry, UK, Technical Report, 2013. Retrieved from eera-es.eu/wp-content/uploads/2016/06/Overview-of-Current-Development-on-Compressed-Air-Energy-Storage_EERA-report-20131.pdf. Accessed 8-May-2020.

D. Mignard. Estimating the capital costs of energy storage technologies for levelling the output of renewable energy sources. *International Journal of Environmental Studies: Turbulent Energy*, 71(6):796-803, 2014. doi:10.1080/00207233.2014.967103.

J. Moore and B. Shabani. A Critical Study of Stationary Energy Storage Polices in Australia in an International Context: The Role of Hydrogen and Battery Technologies. *Energies*, 9(9):674, 2016. doi:10.3390/en9090674.

RWE Power AG. *ADELE - Adiabatic Compressed-Air Energy Storage for Electricity Supply*. RWE Power AG, Report, 2010.

H.I.H. Saravanamuttoo, G.F.C. Rogers, H. Cohen, P.V. Straznicky, and A.R. Nix. *Gas turbine theory* (7th ed.), page 59. Upper Saddle River: Pearson. 2017. ISBN: 9781292093093.

A. Sciacovelli, Y. Li, H. Chen, Y. Wu, J. Wang, S. Garvey, and Y. Ding. Dynamic simulation of Adiabatic Compressed Air Energy Storage (A-CAES) plant with integrated thermal storage – Link between components performance and plant performance. *Applied energy*, 185:16-28, 2017. doi:10.1016/j.apenergy.2016.10.058.

SIEMENS. *Datum Calculator*. v LC, F1.1, B1.1, 2020. Retrieved from http://cirs.dresser-rand.com/turbocalc/Index.html. Accessed 4-April-2020.

D. Wolf and M. Budt. LTA-CAES – A low-temperature approach to Adiabatic Compressed Air Energy Storage. *Applied energy*, 125:158-164, 2014. doi:10.1016/j.apenergy.2014.03.013.

K. Yang, Y. Zhang, X. Li, and J. Xu. Theoretical evaluation on the impact of heat exchanger in Advanced Adiabatic Compressed Air Energy Storage system. *Energy Conversion and Management*, 86:1031-1044, 2014. doi:10.1016/j.enconman.2014.06.062.

O.A. Øvrebø. *Opp 43 prosent: Enda en rekord for norsk vindkraft*, 6-March 2020. Retrieved from energiogklima.no/nyhet/datakilder/status-for-vindkraft-i-norge/. Accessed 10-May-2020.

Sensitivity Analysis of Optimised Large Scale District Heating Heat Pump Concepts

Signe Thomasen[1]   Kim Sørensen[1]   Carsten Bojesen[1]   Kasper Vinther[2]

[1]Department of Energy Technology, Aalborg University, Denmark `sthoma15@student.aau.dk`
[2]Added Values P/S, Denmark, `kvi@AddedValues.eu`

Abstract

This paper investigates the sensitivity on choice of heat pump concept to uncertainties and variations in boundary conditions for large scale heat pumps with a Danish district heating system as case example. A performance map-based heat pump model is used to evaluate performance of different heat pump concepts. The heat pump concepts are optimised in terms of concept design and prepared for evaluation of feedback control parameters and on/off strategies, which calls for a nonlinear optimiser such as genetic algorithms. The optimisation considers the operational price of the concept, regarding Coefficient Of Performance (COP), heating capacity and investment cost and is followed by investigations of the possibility for providing ancillary services. A sensitivity analysis is carried out to investigate whether, which and how variations in cost function weighting affects the optimisation and consequently the choice of heat pump concept. The results indicate a very robust optimised concept.

Keywords: heat pump modelling, large scale heat pumps, optimisation

1 Introduction

By 2050, Denmark should be completely independent of fossil fuels. One essential field in the solution of this is the synergy between the power- and heating grid. With the district heating sector having great potential to be included in solutions across energy sectors, large scale heat pumps have recently received increased interest since it in nature provides a power-to-heat conversion and since tax-reductions and potential for ancillary services likewise suggest large scale heat pumps as a favourable part of the solution (Kortegaard Støchkel et al., 2017). However, focusing on the Danish district heating system, the experience in implementing large scale heat pumps is very limited.

From (Rambøll et al., 2019) it was found that the heating capacity of most large scale units currently in operation in Denmark are in the order of 0.2-10 MW_q. As have been presented in the planned projects in (Fjernvarme Fyn and Odense Kommune, 2020) and (DIN Forsyning and Rambøll, 2020), heat pump units of significantly higher capacity (up to 100 MWq) is

expected to be introduced in the Danish district heating system. A few examples of heat pump units with this high capacities can be found in Norway, (Hoffmann and Forbes Pearson, 2011) and Sweden (Friotherm AG, 2017), and common to these are that the desired capacity is achieved by a combination of a number of smaller heat pump units. There is however no unique formula on how to combine the units. As was stated in (Rambøll, 2019) and in (Meesenburg et al., 2018), especially when bearing the potential for ancillary services in mind, the design of the heat pump concept (i.e. the internal connections and the size of the individual units), should be considered carefully, which due to the lack of experience in ancillary services was found to be challenging.

(Rambøll, 2019) has additionally been conducting some tests of a 0.25 MW_e ammonia heat pump to investigate the potential for providing ancillary services. With the requirements for ancillary services valid for the eastern part of Denmark, it was found that when using the power consumption from the heat pump as set point for control, the response time improved significantly indicating that heat pumps can be applied for providing ancillary services. The experience is however still very limited, and a general approach for designing the large scale heat pump concepts has to the best of the author's knowledge, not yet been presented.

This paper presents a general approach for optimisation of heat pump concepts in future energy systems. The purpose is to develop and demonstrate a heat pump optimisation tool in order to locate the parameters, which dominates the choice of heat pump concept with regard to the number of heat pumps, the size of the heat pumps and the connections between them.

2 Method

To demonstrate the idea of the simulation and optimisation approach, an ongoing project in Esbjerg, Denmark, is used as case example. With boundary conditions and heat pump requirements set by the case, the modelling approach and the simulation- and optimisation setup was developed in three steps, which will be presented in the following sections as:

- A general representation of the heat pump performance.

Table 1. The definition of the seven categories representing an example of the annual operation of the heat pump concept illustrated in Figure 1. Heat load corresponds to the relative amount of heat delivered in each category (in MWh$_q$) on a yearly basis i.a. category 1 is responsible for 2 % of the total amount of heat delivered during a year.

Category	T_{sea} [°C]	$T_{dh,out}$ [°C]	Heat load [%]	Ref. load [%]	α_i [DKK/year]	β_i [DKK/year]
C1	>10	50-55	2	59.8	638 385.0	82 839
C2		60-70	23	86.4	9 077 708.9	
C3	6-10	50-55	14	59.4	5 681 650.3	406 225
C4		60-70	11	69.2	4 600 325.3	
C5	2-6	60-70	39	89.9	17 502 111.2	9 658 168
C6		50-55	3	90.6	1 153 986.	
C7	<2	50-55	9	95.2	4 132 657.0	5 164 456

- A dynamic model.

- A suitable heat pump concept optimisation.

2.1 Case study: Esbjerg District Heating

In Esbjerg, Denmark, a 460 MW$_q$ coal fired heat- and power plant is set to close in March 2023. A considerable amount of heat pumps are included in the road-map for installation of the new heating production. A 50 MW$_q$ sea water based heat pump unit is among other things expected to operate along with wood chip boilers, electric boilers and gas boilers (DIN Forsyning and Rambøll, 2020).

As a part of the initial study of a possible solution, the 50 MW$_q$ is delivered from a heat pump concept consisting of four 12.5 MW$_q$ sea water heat pump units connected internally as seen in Figure 1.
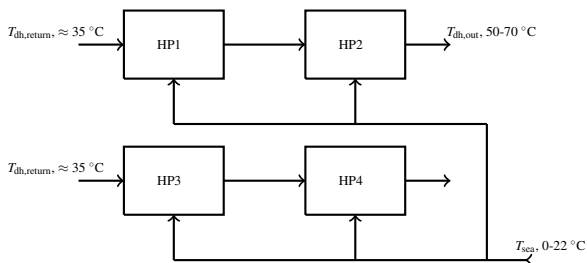
Figure 1. Reference heat pump concept with four 12.5 MW$_q$ heat pumps with two parallel strings each with two heat pumps in series.

The heat pump concept in Figure 1 is referred to as the reference concept. Data on an example of the expected operation of this concept is applied for the heat pump concept optimisation. Divided into seven categories, the annual operation can be considered as seen in Table 1. The categories are chosen instead of the full annual operational profile to smooth out the build in bias towards the reference concept to which the operating profile is optimised. The heat demand presented in Table 1 forms the operating conditions for the simulations conducted. Notice, that these categories change, if the heat pumps are combined with e.g. wind power or combined heat and power plants. The heat pump system used for simulation is presented in the following sections.

2.2 Model

The heat pump model and the simulation model is formulated in the multi-domain, object oriented modelling language, Modelica (Modelica Association, 2019) and implemented in Dymola (Dassault Systémes, 2019). The heat pump system makes use of pipe, flow source/sink and sensor models found in the Buildings library (Wetter et al., 2014). Using the Modelica language implies a high degree of reusability and flexibility in the modeling and does further allow for easily exporting the model as a functional mockup unit (FMI) and further simulating the model in e.g. Python through the functional mockup interface (FMI) standard.

2.2.1 Heat pump model

As suggested in (Dott et al., 2013) and similarly applied in the energy simulation software EnergyPlus (Department of energy, 2019), a typical approach for performance evaluation of heat pumps is by a *performance map*. A common method for applying the performance map approach is to use data points either to interpolate between or by using an equation fit to create a polynomial plane. Inspired by this, the performance of the heat pump model is described by two equations developed from data obtained using the product selection software, GEA RTSelect, provided by GEA (GEA Group, 2019). The performance map describes the heat output, $\dot{Q}_h$ and the Coefficient of Performance, COP of a single heat pump unit. The performance degradation due to part load operation is included in the performance map by introducing both the desired outlet temperature of the district heating water, $T_{dh,out}$ [K], the evaporation temperature, T_{eva} [K] and load, L [%] as variables for the equation fit.

By stepwise including the three variables to the fit of the performance data, Equation 1 and Equation 2 was created using R (R Foundation, 2019).

$$\dot{Q}_h = \theta_0 + \theta_1 L + \theta_2 L^2 + \theta_3 T_{eva} + \theta_4 T_{dh,out} + \theta_5 L T_{dh,out} + \theta_6 L T_{eva} \tag{1}$$

With θ_i and γ_i being the regression coefficients found in Table 2 and Table 3. Applying Equation 1 to test data showed a maximum deviation of $\approx$ 5 %. A similar test on the validity of Equation 2 shows a maximum deviation of $\approx$ 7 %. It should additionally be noticed, that the

Table 2. Values of the regression coefficients, θ for the equation fit of the heat output.

θ_0	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
- 241 000	- 1 063 000	70.96	-1 667	10 030	4001	306.7

map (Equation 1 and Equation 2) is only guaranteed valid within the range of the data. This corresponds to a district heating outlet temperature between 45 °C and 80 °C and a sea water temperature between 2 °C and 17 °C.

$$
\begin{aligned}
\text{COP} = {} & (\gamma_0 - \gamma_1 \exp(-\gamma_2 L)) + \gamma_3 T_{\text{eva}} + \gamma_4 T_{\text{eva}}^2 \\
& + \gamma_5 (\gamma_0 - \gamma_1 \exp(-\gamma_2 L)) T_{\text{eva}} + \gamma_6 T_{\text{dh,out}} + \gamma_7 T_{\text{dh,out}}^2 \quad (2) \\
& + \gamma_8 (\gamma_0 - \gamma_1 \exp(-\gamma_2 L)) T_{\text{dh,out}} + \gamma_9 T_{\text{eva}} T_{\text{dh,out}}
\end{aligned}
$$

At a constant evaporation temperature of 5 °C, the surface of Equation 2 is seen in Figure 2. Equation 1 and Equation 2 form the basis of the heat pump model in Dymola.

The heat pump model consists mainly of three components: a *heat moving component*, a pipe representing the condenser and a pipe representing the evaporator, as seen in Figure 3.
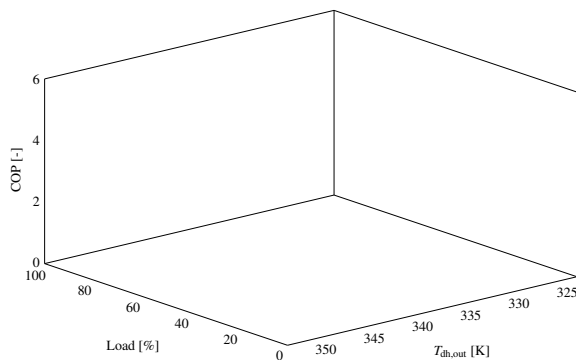
Figure 2. Surface plot of the regression equation for COP, with a evaporation temperature of 5 °C.

The *heat moving component* includes the performance map described by Equation 1 and Equation 2 and the correlations presented in Equation 3 and Equation 4.

$$
\dot{Q}_{\text{c}} = P(\text{COP} - 1) \quad (3)
$$

$$
P = \dot{Q}_{\text{h}} - \dot{Q}_{\text{c}} \quad (4)
$$

The combination of Equation 1 to Equation 4 represents the amount of heat added to the condenser pipe, the amount of heat removed from the evaporator pipe and the electricity consumed by the heat pump unit. The two pipes are found from the buildings library and allow for adding/extracting a prescribed heat flow to the fluid flow in the pipe. Using the buildings library, (Wetter et al., 2014), additionally provides a simplified medium model (assuming constant density) which improves the simulation time.

The performance map approach and the chosen Modelica formulation do however not include any information on the dynamic behaviour of the heat pump unit. For the compressor, the dynamic effects are accounted for by limiting the allowable load gradients corresponding to $t_{\text{min-max}} = 120 s$ and $t_{\text{max-min}} = 60 s$. Notice that this is not the actual compressor dynamics, but represents the effect of the dynamics related to the compressor.
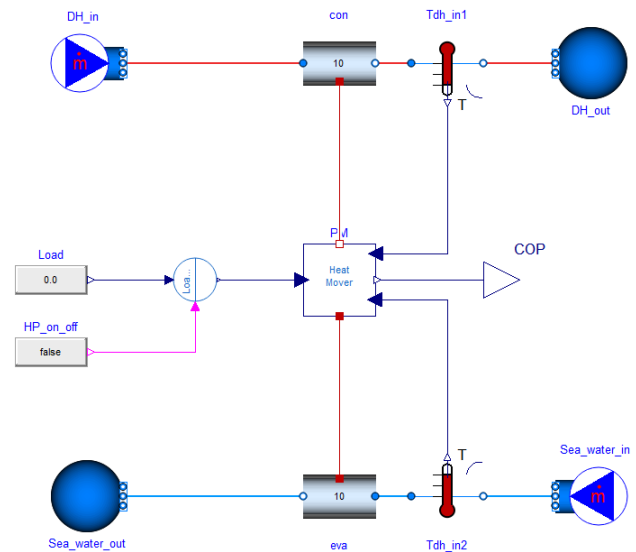
Figure 3. Simplified sketch of the heat pump unit i Dymola. The heat mover contains the performance map equations.

In (Farouk Fardoun et al., 2011) it was however found that the dominating dynamic is related to the heat exchangers and not the compressor. The dynamic in the heat exchangers is accounted for by the volume of the pipes representing the heat exchangers, which is set to $\approx 4 \text{ m}^3$.

2.2.2 Heat Pump System

The heat pump model described in the previous section forms the base of the model used for simulation and following optimisation. The general idea of the simulation model is sketched in Figure 4. Based on the total 50MW_q heat requirement for the case in Esbjerg, a maximum of nine units is assumed. The nine heat pump units can be arranged with maximum three parallel strings each with maximum three heat pumps in series. As appears from Figure 4, each heat pump unit has its own mass flow source. However, the flow in each parallel string is identical, and the individual mass flow sources are introduced to avoid including dynamic effects of heat pumps, which are not included in the concept when performing the optimisation.

Table 3. Values of the regression coefficients, γ for the equation fit of the COP.

γ_0	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7	γ_8	γ_9
- 12.52	3.076	-0.074	0.710	0.001	0.030	-0.474	0.001	-0.024	-0.002
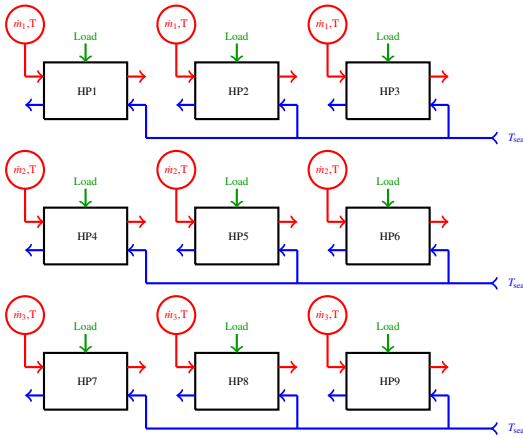
Figure 4. Simplified sketch of the heat pump concept simulation model with nine possible heat pumps with a maximum 3-by-3 configuration.

2.2.3 Control

Depending on whether the heat pump concepts are evaluated according to their COP, heating capacity or dynamic response, different requirements or approaches exists. For evaluation of COP and heating capacity, both load control and temperature control are applied. The temperature control adjusts the mass flow rate of the district heating water to achieve the desired outlet temperature. This control is only applied for the evaluation of COP and heating capacity.

For the load control, two approaches can be used: heat output control or power consumption control. The heat output control is applied for steady state evaluation of the COP and heating capacity, while the power consumption control can be applied for evaluation of dynamic performance. Last mentioned is especially an advantage when considering ancillary services, since this was found to allow for faster regulation (Rambøll, 2019).

3 Heat Pump Concept Optimisation

From the heat pump simulation model in Section 2, the optimisation presented in this paper seeks to find the optimum heat pump concept according to its steady state COP, heat output and the depreciation expense.

3.1 The Concept Optimisation Problem

The vector $\mathbf{u}$ contains the optimisation variables.

$$\mathbf{u} = \{\dot{Q}_{h,n,1}, \dot{Q}_{h,n,2}, \ldots, \dot{Q}_{h,n,9}\}$$

With $\dot{Q}_{h,n,i}$ representing the nominal heating capacity of a heat pump unit, $0 < \dot{Q}_{h,n,i} < 18.5$ MW$_q$. All units with $\dot{Q}_{h,n,i} \leq 2$ are truncated to 0, and is consequently not included in the concept. The annual expenses include the

costs related to COP, the total heating capacity of the concept and the total depreciation expenses. Based on the case in Esbjerg, the expected operation and the expected performance of the reference concept give rise to some expected costs. Changing the reference concept in Figure 1 will in most cases also change its performance, and the investment cost. The concept optimisation problem can be formulated as seen in Equation 5.

$$\begin{aligned} \underset{\mathbf{u}}{\text{minimise}} \quad & \text{COST}_{\text{HP}} \\ \text{subject to} \quad & \dot{x}(t) = f(\mathbf{V}, \mathbf{u}, \mathbf{x}(t)), \\ & \mathbf{G}\mathbf{u} \geq \mathbf{g}, \\ & \mathbf{x}(0) \geq \mathbf{x_0} \end{aligned} \tag{5}$$

With COST_{HP} representing the annual costs for the heat pump, $\mathbf{x} \in \mathbb{R}^q$ being the system state vector and $\mathbf{V} \in \mathbb{R}^m$ the input vector with external boundary conditions. The nonlinear, dynamic model formulated in Dymola, is represented by f, and the constraints in the matrix $\mathbf{G}$ and vector $\mathbf{g}$.

3.2 Micro-Genetic Algorithm

Due to the non-linearity of the optimisation problem, there could be multiple local suprema, and a global optimisation method must consequently be applied. For this purpose, the μGA proposed in (Krishnakumar, 1990), seen in Figure 5 is proposed and the specific parameters are based on suggestions from (Vinther et al., 2017).

Initialisation

Using a population size of $N_s = 5$, individuals are picked randomly from a normal distribution with center defined by the best solution so far and with a standard deviation set to 50 % of the distance between minimum and maximum bounds of the optimisation variables. Additionally, the picked parameter is truncated to be within the min/-max bounds, which means that the odds for picking values close to the current best solution and at the bounds increase.

Automatic restart

To avoid getting stuck at a local optimum, an automatic restart is triggered whenever the individuals are converged. This happens when all pairs in all individuals are within 5 % of their full range. The automatic restart mean that new individuals are picked as presented in the previous section while keeping the best (elite) individual.

Fitness evaluation

The fitness of each individual is calculated using a cost function.

Deterministic Tournament Selection

Two individuals are picked randomly from the population and the most fit of the picked individuals are chosen as a parent for crossover.

Crossover - Gaussian Fitness Based

New individuals (children) are produced from the parents. This is done by randomly picking values from a normal distribution with center closest to the fittest parent and the standard deviation set to 50 % of the distance between the parents.
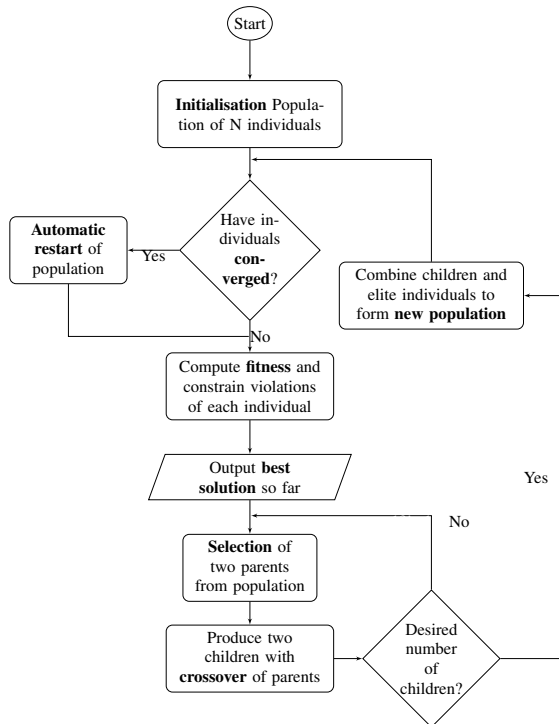
Figure 5. Flow chart of the μGA used for the optimisation of the heat pump concept.

3.3 Cost Function

To evaluate the fitness of different heat pump concepts according to their COP, heating capacity and investment cost, the cost function in Equation 6 is introduced.

$$\text{COST}_{\text{HP}} = \text{COST}_{\text{COP}} + \text{COST}_{\text{Heat}} + \text{COST}_{\text{CAPEX}} \quad (6)$$

The general idea of the cost function in Equation 6 is to represent the difference in annual cost between the reference concept in Figure 1 and the concept in question. Descriptions of each term in Equation 6 (COP, heating capacity and investment cost) are presented in the following sections.

3.3.1 Cost: Coefficient of Performance

Increasing the COP compared to the reference concept leads to an overall reduction in electricity consumption from the heat pumps and vice versa. The steady state COP

is evaluated in each of the seven categories presented in Table 1 as seen in Equation 7.

$$\text{COST}_{\text{COP}} = \sum_{i=1}^{7} \alpha_1 \left(1 - \frac{\text{COP}_i}{\text{COP}_{\text{ref},i}}\right) \quad (7)$$

The unit of the coefficients, α_i, is in DKK/year, and represents the cost penalty or profit achieved from either reductions or improvements in COP, weighted according to the energy content in each category as seen in Table 1. This assures, that the highest profit or penalty is associated with the category with the highest heating demand.

3.3.2 Cost: Heating Capacity

Combining category 1 and 2, 3 and 4, 5 and 6 and keeping category 7 from Table 1 forms the boundary conditions applied for the evaluation of the heating capacity. In each of these categories, the reference concept is capable of delivering 66.7 MW$_q$, 55.8 MW$_q$, 49.3 MW$_q$ and 44.3 MW$_q$ respectively. These maximum capacities are used as a benchmark for the evaluation of the heating capacity. If the capacity of the heat pumps is reduced, other units (gas boiler, wood chip boiler etc.), must compensate leading to increased costs. On the other hand, increasing the capacity is undesirable, meaning that no profit can be achieved by increasing this. The cost function describing the heating capacity is seen in Equation 8.

$$\text{COST}_{\text{heat}} = \sum_{i=1}^{4} \max\left(0, \beta_1 \left(1 - \frac{\dot{Q}_{\text{h},i}}{\dot{Q}_{\text{h,ref},i}}\right)\right) \quad (8)$$

Similar to Equation 7 for COP, the β_i coefficients are weighted according to the heating load in each category. The values of α_i and β_i are found in Table 1.

3.3.3 Cost: Investment and Depreciation Expense

The reference price and the assumed effects of economy of scale are based on numbers presented by (Danish Energy Agency and Energinet, 2016). This mean, that the reference price is based on a 4 MW$_q$ heat pump with a cost of 4.95 MDKK/MWq. The effect of economy of scale is set to 0.8. Based on this, the total investment cost is described by Equation 9

$$\text{COST}_{\text{inv}} = \sum_{i=1}^{9} 6e6 \, \dot{Q}_{\text{h,n},i}^{0.848} \quad (9)$$

Expressing the investment cost relative to the reference case and in terms of depreciation expense results in Equation 10. A 20 year annuity with an interest rate of 1.8 % is assumed.

$$\text{COST}_{\text{CAPEX}} = (0.06 \, \text{COST}_{\text{inv}} - 0.06 \, \text{COST}_{\text{inv,ref}}) \quad (10)$$

4 Results and Discussion

4.1 Optimisation

The optimisation is initially conducted using the cost function as it is described in the previous section, with a maximum allowable heat pump unit size of 18.5 MW$_q$.

4.1.1 Convergence

One challenge when optimising a non-linear system, is to ensure that the suggested optimum is a global optimum. The optimisation allows the modeller to provide some guesses on heat pump concepts for the initial generation. However, if a global optimum is found, the end result both with and without a guess on a heat pump concept should be consistent. For a 500 generation run this was tested and the result is presented in Figure 6.
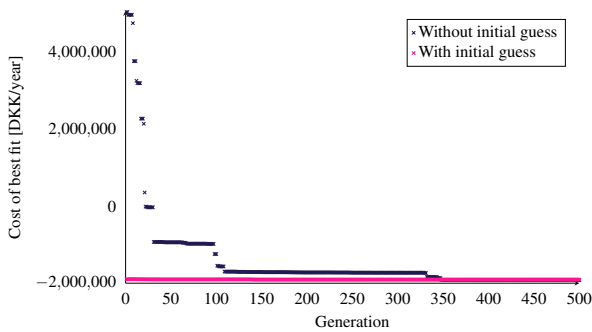
Figure 6. Convergenve plot showing the cost of the best fit during 500 generation test runs with and without initial guesses.

Figure 6 shows that the end results are consistent when running 500 generations with and without initial guesses. This indicates that a global optimum has been reached. Figure 6 also indicates, that the number of required generations reduces significantly, when the initial guess is close to the optimum. However, care must be taken for a low number of generations as this might lead to a local optimum.

4.1.2 Optimised Heat Pump Concept

A sketch of the optimised heat pump concept is seen in Figure 7.
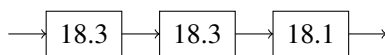
Figure 7. A sketch of the optimised heat pump concept. The numbers inside the each heat pump box represent the chosen nominal capacity of the heat pump.

As seen from Figure 7, the optimum heat pump concept at the given conditions consists of three heat pump units in series each with a capacity of ≈ 18.2 MW$_q$. Notice, that these heating capacities sum to ≈ 54.5 MW$_q$, which is due to the different nominal conditions assumed for the performance map. The costs for the optimised concept are allocated as seen in Table 4.

Table 4. Cost allocation for the optimised heat pump concept.

Cost [DKK/year]	Optimised concept
Cost$_{COP}$	- 1 385 935
Cost$_{heat}$	1417
Cost$_{Invest}$	- 501 896
Total Cost	-1 886 414

Compared to the reference the results show, that by increasing the number of heat pumps in series to three instead of two, the COP improves. This is consistent with expectations. At the same time, the cost pr. MW$_q$ reduces due to the increased unit size and the reduced number of units.

4.2 Sensitivity Study

The results of the optimisation provides the best possible concept in terms of COP, heating capacity and investment cost. However, from the investors point of view the robustness of the optimised concept is not insignificant. To demonstrate the capability of the simulation model- and optimisation setup, a sensitivity on 3 parameters related to the cost function will be presented in the following sections.

4.2.1 Variations in Costs for Electricity

The penalty/profit related to reductions or improvements in COP is based on the expected costs for electricity. Consequently, changes in the expected cost for electricity naturally leads to changes in the total annual costs for the heat pump concept. A significant change of $\pm$ 40 % was considered initially. The results of the two optimisations are presented in Figure 8.

(a) Elspot + 40 %　　　　**(b)** Elspot - 40 %
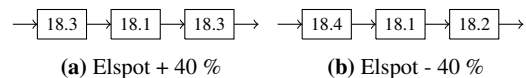
Figure 8. A sketch of the optimised heat pump concept, when the estimations of the expected elspot price is reduced by - 40 % (fig. a) and increased by + 40 % (fig. b).

The optimum concepts for both an increase and a reduction in costs for electricity are, as seen in Figure 8, similar to the concept found from the original optimisation. The costs are allocated as seen in Table 5.

Table 5. Cost allocation for optimisation with the cost for electricity increased by 40 % and reduced by 40 %.

Cost [DKK/year]	+ 40 %	- 40 %
Cost$_{COP}$	- 1 728 332	- 1 078 343
Cost$_{heat}$	2 889	1 728
Cost$_{Invest}$	- 503 097	- 502 245
Total Cost	- 2 228 542	- 1 578 865

Since the concepts are similar to the concept found from the original optimisation, the costs regarding heating capacity and investment are similar as well. As expected, an increase in costs for electricity increases the profit achieved by improving the COP, whereas reduced costs for electricity reduces the profit achieved from improving the COP. With three large units, the investment costs are at a minimum due to the effects of economy of scale, and since the COP is maximised by connecting the three large heat pump units in series, the result is consistent with expectations.

4.2.2 Adjusting the Benchmark for Investment Cost

The investment costs in the initial optimisation is based on the numbers presented in (Danish Energy Agency and Energinet, 2016). However, a lower cost estimation and a higher cost estimation is presented as well. Similar to the electricity costs, adjusting the benchmark for the investment costs does not change the design of the optimised heat pump concept, meaning that the the result of the optimisation again suggest three serial coupled heat pumps of ≈ 18.2 MW$_q$. The costs are allocated as seen in Table 6.

Table 6. Cost allocation for optimisation with the low cost estimation benchmark and with the high cost estimation benchmark (Danish Energy Agency and Energinet, 2016).

Cost [DKK/year]	Low benchmark	High benchmark
Cost$_{COP}$	- 1 403 569	- 1409 097
Cost$_{heat}$	1 306	74 784
Cost$_{Invest}$	- 336 398	- 922 706
Total Cost	- 1 738 661	- 2 257 019

Observing the investment costs, the low benchmark produces a reduction in the effect of investment cost whereas the high benchmark increases the profit from choosing the concept with the lowest possible investment costs. On the other hand, the increased investment reference price implies that the optimisation reduces the capacity to a minimum in order to achieve the highest investment cost profit.

4.2.3 Changes in Effects of Economy of Scale

The assumed economy of scale-effect of 0.8 is as the last parameter varied. Firstly by increasing the effect to 0.5, meaning that a doubling of the heat pump capacity only give rise to a 50 % cost increase. Secondly the effect of economy of scale was removed completely. The optimum concepts under these conditions are seen in Figure 9.

(a) EOS = 50 %   **(b)** No effects of EOS
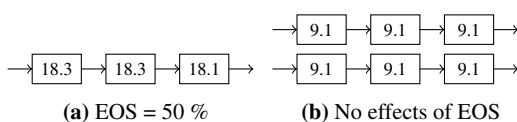
Figure 9. A sketch of the optimised heat pump concept, when the effects of economy of scale are incresed to 50 % (fig. a) or removed completely (fig. b).

The noticeable result from this optimisation, is that removing the effects of economy of scale, does, contrary to all other sensitivities conducted, produce a different optimised heat pump concept. The reason for this becomes more apparent when considering the cost allocations presented in Table 7.

Table 7. Cost allocation for optimisation with increased effects of economy of scale (0.5) and with no effects of economy of scale.

Cost [DKK/year]	EOS = 0.5	EOS = 1
Cost$_{COP}$	- 1 403 654	- 1 721 945
Cost$_{heat}$	1 703	52 540
Cost$_{Invest}$	- 777 417	38 286
Total Cost	- 2 179 368	- 1 631 119

By having two parallel strings instead of one as suggested in the original optimisation, the profit from COP increases from $\approx$ 1.4 million DKK/year to $\approx$ 1.7 million DKK/year. The reason for this can be found in the chosen control strategy, which ensures that whenever more than one string is present in the heat pump concept, the second string, third string etc. will remain inactivated until the first parallel string reaches 100 % load. The apparent consequence of this is that the COP during the first categories increases due to increased load-levels caused by the reduced activated capacity, when the second string has not yet been activated. On the other hand, removing the effect of EOS makes it impossible to achieve any profits in terms of investment costs and for that reason, the total annual savings are slightly less than for the original optimisation.

One noticeable potential achieved from the combination of the dynamic model in Dymola and the chosen optimisation algorithm is that it allows to include a wide number of parameters in the optimisation. In this paper, only the steady state COP, heating capacity and investment costs of the heat pump concept are included. However, the setup has the potential to include Dynamic effects as e.g. control parameters and ancillary services as well. Parameters as maintenance and ensuring redundancy would most likely also affect the final choice of heat pump concept. Last mentioned is especially interesting regarding the concept suggested by the optimisation, since it has only a single parallel string. A consequence of this is that in order to ensure redundancy, extra piping etc. for bypass would be necessary, which in the end increases the investment costs. Additionally, requirements and potential limitations defined by the heat pump suppliers could furthermore influence the result of the optimisation. These effects could, similar to dynamic effects and control strategies be included in the optimisation if it is quantified.

To demonstrate, that other parameters such as the potential for providing ancillary services might affect the optimisation results, a small capacity test on the reference concept and the optimised concept was conducted. Assuming that during ancillary services, the maximum capacity provided corresponds to one parallel heat pump string running at 10 % load. For the reference concept this corresponds to a maximum regulation capacity of 11.1 MW$_e$ while the optimised concept has a maximum of 10.4 MW$_e$. The question is whether this difference in capacity for bids affects the annual costs for the concept. Assuming the following availability payments (Energinet, 2019), the difference in annual profit from availability payments are found in Figure 10:

- FCR$_{up}$ = 276 DKK/MW$_e$/h

- FCR$_{down}$ = 26 DKK/MW$_e$/h

- aFRR = 150 DKK/MW$_e$/h
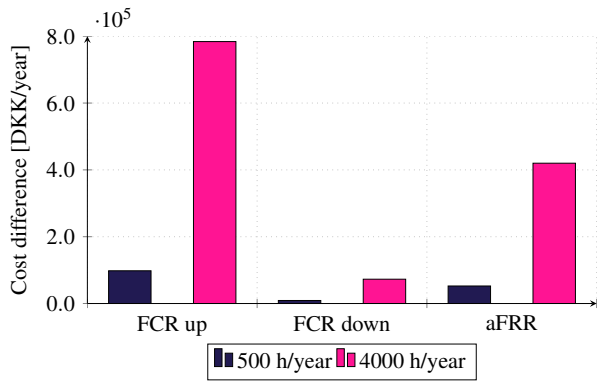
Figure 10. Lost availability payment from ancillary services relative to reference concept for 500 h/year and 4000 h/year respectively.

Figure 10 illustrates first of all that the difference in potential availability payment depends on the amount of hours won. Secondly, it can be seen that if assuming 4000 availability hours per year, the difference in maximum capacity implies a significant profit increase for the reference case.

5 Conclusion

A heat pump concept optimisation setup was developed and demonstrated based on an ongoing project in Esbjerg. The combination of a dynamic model formulated in Dymola and the micro-genetic algorithm formulated in Python was found to be a promising tool for heat pump concept optimisation, especially due to the potential to include everything from the relatively simple steady state COP evaluations presented in this paper, to more complex dynamic effects as control parameters and ancillary services.

The sensitivity analysis indicated that with the optimisation variables and cost functions presented in this study, the optimised heat pump concept is very robust to changes in both costs for electricity and the basis of the calculation of the investment costs. Removing the effects of economy of scale does however change the optimum concept from a single heat pump string with three large scale heat pump to a concept with two parallel strings each with three heat pumps in series.

References

Danish Energy Agency and Energinet. Technology Data - Generation of Electricity and District heating. Technical report, August 2016.

Dassault Systémes. Dymola systems engineering, 2019. URL https://www.3ds.com/products-services/catia/products/dymola/.

U.S. Department of energy. Engineering Reference. *EnergyPlus^{TM} Version 9.1.0 Documentation*, (9.1.0): 1748, 2019.

DIN Forsyning and Rambøll. Din Forsyning A/S projektforslag havvands og luft varmepumpeanlæg samt fliskedelanlæg. Projektforslag fra DIN Forsyning A/S, Esbjerg Kommune, 2020.

Ralf Dott, Thomas Afjei, Andreas Genkinger, Antoine Dalibard, Dani Carbonell, Ricard Consul, Andreas Heinz, Michel Haller, Andreas Witzig, Jorge Facão, and Peter Pärisch. A

technical report of subtask C Report C2 Part C – Final Draft. 2013:97, 2013.

Energinet. Årlig statistik over rådighedsbetaling for reserver i øst- og vestdanmark, samt statistik over omfanget af specialregulering inkl. nøgletal for 2019, 2019.

Farouk Fardoun, Oussama Ibrahim, and Assaad Zoughaib. Dynamic modeling of an air source heat pump water heater. May 2011.

Fjernvarme Fyn and Odense Kommune. Udfasning af kul på fynsværket. Technical report, Fjernvarme Fyn & Odense Kommune, 2020.

Friotherm AG. Värtan Ropsten - the largest sea water heat pump faclity worldwide, with 6 Unitop 50FY and 180 MW total capacity, 2017.

GEA Group. GEA RTSelect software, 2019.

Kenneth Hoffmann and David Forbes Pearson. Ammonia Heat Pumps for District Heating in Norway – a case study. page 8, 2011.

Hanne Kortegaard Støchkel, Bjarke Lava Paaske, and Kim S. Clausen. Drejebog til store varmepumpeprojekter i fjernvarmesystemet. Technical report, Grøn Energi, PlanEnergi, DFP, December 2017.

Kalmanje Krishnakumar. Mircro-Genetic Algortihms For Stationary And Non-Stationary Function Optimisation. page 9, 1990. doi:10.1117/12.969927.

Wiebke Meesenburg, Torben Ommen, and Brian Elmegaard. Dynamic exergoeconomic analysis of a heat pump system used for ancillary services in an integrated energy system. page 12, 2018.

Modelica Association. Modelica association. modelica and the modelica standard library., 2019. URL https://modelica.org/.

R Foundation. The r project for statistical computing, 2019. URL https://www.r-project.org.

Rambøll. Ancillary Services From New Techonologies. Technical report, Rambøll, 2019.

Rambøll, Dansk Fjernvarme, and Grøn Energi. Store Varmepumper i Fjernvarmen Driftserfaringer. Technical report, October 2019.

K. Vinther, Rene J. Nielsen, Palle Andersen, and Jan D. Bendtsen. Optimization of interconnected absorption cycle heat pumps with micro-genetic algorithms. *Journal of Process Control*, 53:26–36, May 2017. ISSN 09591524.

Michael Wetter, Wangda Zuo, Thierry S. Nouidui, and Xiufeng Pang. Modelica Buildings library. *Journal of Building Performance Simulation*, (4), July 2014. doi:10.1080/19401493.2013.765506.

Simulation of heat recovery from data centers using heat pumps

Mohammad Sharfuddin Lars Erik Øi

Department of Process, Energy and Environmental Technology, University of South-Eastern Norway

`lars.oi@usn.no`

Abstract

Digitalization has influenced the rapid growth of data centers around the world. The advancement of IT and telecommunication also played a vital role in this expansion of data centers. Data centers facilitate the storage and access of data when required. Electric power is the main energy input and heat is the main energy output from the data center. This work is about the utilization of the excess heat which is the by-product of data center operation. Possible ways to utilize waste heat from data centers have been evaluated. To connect the heat from data centers to a district heating network, a heat pump might be necessary to increase the temperature of the heat. Simulations were performed at varying conditions in Aspen HYSYS to evaluate the waste heat utilization. The economic potential for different conditions and different heat recovery solutions are also evaluated.

Keywords: data center, district heating, heat recovery, heat pump, Aspen HYSYS.

1 Introduction

Data centers have become an indispensable part of the modern digitalized world. Digitalization demands the storage of constantly and rapidly expanding data around the globe, for which the number of data centers is ever-increasing. Apart from digitalization, speedy wireless networks, growing demand for cloud computing have added the crying needs of data centers. The U.S. Environment Protection Agency defines a data center as (Geng, 2014): *"Primarily electronic equipment used for data processing (servers), data storage (storage equipment), and communications (network equipment). Collectively, these equipment processes, stores, and transmits digital information".*

Data centers are run by electricity and the functioning of different equipment release heat. So, all the electricity input is converted to heat. Studies show that the electricity requirement for data center has increased from about 1.3% of the world's total electricity consumption in 2010 to 2% in 2018 and with this pace, it will reach up to 13% in 2030 (Oltmanns *et al.*, 2020).

However, the energy source which is still dependent on the fossil fuels are gradually decreasing because of the ever-growing consumption. So, the dire need for renewable energy sources is beyond description. The rejected or output heat from data centers could be a useful source of renewable energy. So, the waste heat utilization from data centers has become one of the prime researches for the scientists and data center operator to make data centers energy efficient and economically sound.

Data centers reject a vast amount of heat which is the conversion of electricity. For the proper and reliable functioning of the data center cooling down of different IT equipment is essential. 40% of the total energy consumption in a DC can be spent in the cooling system (Capozzoli and Primiceri, 2015). Moreover, excess or rejected heat from a DC can be regarded as lost energy. To make data centers more energy-efficient these lost energies should be utilized. The utilization will be economically profitable also. Furthermore, the cold climate of Nordic countries makes it easy for data centers to provide cooling energy. Besides, the high demand for heat in these countries makes it more convenient to utilize the waste heat. Thus, the necessity of waste heat utilization from data centers arises. The work will also focus on the possible utilization of waste heat using a heat pump in the district heating facility.

2 Literature Review

2.1 General literature on energy recovery from data centers

Ebrahimi *et al.* (2014) investigated different waste heat recovery technologies from the data center. They suggested that district heating is a common low-quality waste heat recovery technique which is also economically and ecologically sound. Liquid-cooled servers are more compatible with the higher waste recovery temperatures. Liquid-cooled servers can provide waste heat of up to 50-60℃ (Ebrahimi *et al.*, 2014) that can be applied to district heating over a long area. This waste heat recovery technique is economically profitable as it can earn a revenue stream for data center operators.

The next option found by Ebrahimi *et al.* (2014) was the heating of water in a thermal Rankine cycle. The waste heat cannot fully replace the boiler but can be used to preheat boiler feedwater. So, the consumption of fossil fuel and pollution can be decreased to some extent. Moreover, they suggested on-chip two-phase cooled data centers to utilize most of this technology. The sale of heat to the power plant and carbon offsets can produce substantial income.

Absorption cooling is another choice for utilizing data center waste heat as studied by Ebrahimi *et al.* (2014) Absorption refrigeration systems can function with generator temperatures of 70-90ºC which could be supplied by the waste heat from a water-cooled and two-phase cooled data center. An air-cooled data center is not viable for the technology. This technique can minimize the load on data center CRAC (computer room air conditioning) systems by producing chilled water for cooling and thus become economically profitable.

Ebrahimi *et al.* (2014) also proposed that organic Rankine cycle (ORC), multiple-effect distillation (MED), direct power generation like piezoelectric and thermoelectric, biomass co-location are the possible techniques that can be useful to utilize the low-grade waste heat from data centers.

Oltmanns *et al.* (2020) proposed a new cooling concept which TU Darmstadt will employ in the next generation of the current air-cooled servers with water-cooled rear doors. The new data center will use direct hot-water cooling for the high-performance computer, providing heat at 45ºC. The waste heat will be utilized for heating the university's campus Lichtwiese. They suggested two ideas, either heat integration in the return line of the district heating network or utilizing it locally in buildings situated near the data center. The project showed that 20-50% of the waste heat rejected by the high-performance computer can be utilized in the heating sector. A significant reduction of CO_2 emission can also be achieved through the project.

Oró *et al.* (2018) studied a liquid-cooled on-chip server numerically for a case study of utilizing the waste heat for an indoor swimming pool heating. For the most suitable solution, the data center operator decreases its operational costs and produces surplus income by selling the excess heat, obtaining a net present value after 15 years of 330,000 €. Besides the operational cost of the indoor swimming pool was reduced by 18%. The case study was implemented for the assessment of Barcelona's indoor swimming pools.

2.2 Possible temperatures in cooling principle in data centers

For the efficient and proper utilization of excess heat from the data centers determining the temperature of the cooling system is not only very essential but also very sensitive. Depending on the temperature range the quality of the heat will be evaluated. The investigation is not a very easy task rather it has been a matter of argument.

ASHRAE Technical Committee 9.9 has done a significant job to determine the favorable environment and temperature range for data centers. This is a common thermal guideline. ASHRAE (2011) recommended that the data center's equipment should maintain the temperature range between 18ºC and 27ºC to fit the manufacturer's provided criteria. The Technical Committee also classified the data center based on their temperature range. For the A1 data center, the temperature range was 15ºC to 32ºC, for the A2 category the range was set to 10ºC to 35ºC. For class A3 and A4 data center they increased the temperature range by 5ºC to 40ºC and 5ºC to 45ºC, respectively.

Oltmanns *et al.* (2020) studied that the high cooling inlet temperatures of up to 60ºC for water-cooled data center allow the possibilities for better waste heat utilization.

According to Patel (2003) for an efficient air-cooling system, the cold air should be supplied at 25ºC and exhaust air should exit the room and come back to CRAC at 40ºC.

Ebrahimi *et al.* (2014) suggested that the optimum temperature range to utilize the waste heat in the air-cooled data center at rack exit is 30-40ºC while for the chiller water return the suitable temperature range is 16-18ºC.

Brunschwiler *et al.* (2009) found that the inlet water temperature can be 60ºC to keep the junction temperature under 85ºC. For this criterion, the maximum inlet temperature could be as high as 75ºC.

Sharma *et al.* (2012) depicted that to recover maximum waste heat the suitable inlet temperature can be in the range 40-40.7ºC. They suggested microprocessor junction temperature can be a maximum of 90ºC.

3 Process Description

Heat pump technology gives an effective and long-lasting solution for both heating and cooling applications. A conventional heat pump is a system working on the compression refrigeration cycle powered by either mechanical energy or electricity (Øi and Tirados, 2015). In data center cooling for both air-cooled and liquid-cooled process heat pump is an essential part that regulates the cooling medium's temperature. Typical refrigerants used in heat pumps are ammonia and chlorinated or fluorinated hydrocarbons electricity (Øi and Tirados, 2015).

In the refrigeration cycle, the refrigerant circulates due to temperature and pressure difference between the components. The four main components of

a refrigeration cycle are the compressor, condenser, expansion valve, and evaporator. Figure 1 depicts the mechanical compression refrigeration cycle of a traditional heat pump. The red lines represent high pressure and temperature and the blue line indicates low pressure and temperature of the refrigerant. The cooling effect is produced by the cold liquid refrigerant in the evaporator. A mixture of vapor and liquid phased refrigerant goes into the evaporator where the vaporization of liquid provides the cooling effect before leaving the evaporator. The vapor refrigerant is sucked by the compressor where it gains high pressure and becomes superheated. The output from the compressor then enters the condenser. In the condenser, the vapor refrigerant is cooled and condensed to a saturated liquid. Heat is released from the refrigerant to the ambient (Smith, 2005).
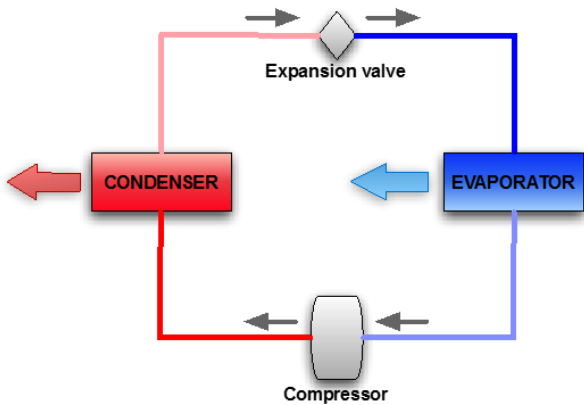
Figure 1. Schematic diagram of a heat pump's mechanical compression cycle (Øi and Tirados, 2015).

The liquid refrigerant then enters the expansion device typically an expansion valve where it is expanded to lower pressure. The liquid refrigerant is partially vaporized due to the expansion process giving a cooling effect in the refrigeration cycle (Smith, 2005).

The efficiency of a heat pump is measured by the coefficient of performance (COP). It is the ratio of the heat delivered or supplied at high temperature to the required power. Equation 1 represents the COP of the heat pump (Øi and Tirados, 2015).

$$COP = \frac{Q_C}{W} \qquad (1)$$
$$W = Q_C - Q_E \qquad (2)$$

In Equation 2 Q_C is the amount of heat output from the condenser, Q_E is the amount of heat input from the evaporator, and W is the power required in the compressor. When there is no heat loss the work added in the refrigeration cycle is equal to the difference between heat output and heat input (Øi and Tirados, 2015).

Typical COP values calculated in the work of Øi and Tirados (2015) are between 3 and 10, dependent on the difference between the delivery and output temperature.

4 Process Simulation, Results and Discussion

4.1 Simulation setup in Aspen HYSYS

For calculation and simulation first, the simulation was set up in the Aspen HYSYS. Version 10 of Aspen HYSYS was used for simulation. In the component lists two pure components described. The components are pure water and pure Refrig-22(R-22). R-22 was selected as the refrigerant medium and water which will be supplied for the cooling process in the data center was selected. After that, in the fluid packages, Peng-Robinson (PR) package was selected which is the most common and efficient package for HYSYS simulation. The default parameters for the package was used by Aspen HYSYS. Then the units of the heat pump which are evaporator, condenser, compressor, and expansion valve, were defined for the simulation with relevant streams.

4.2 The energy required calculation from Aspen HYSYS

One of the important tasks of the work is to perform the calculation in Aspen HYSYS. Two alternatives were selected for the simulation in the Aspen HYSYS. The setup condition for the alternatives are shown in Table 1 and Table 2.

Table 1. Aspen HYSYS input condition for alternative 1

Stream name	Water 1	Water 6
Temperature (ºC)	65	70
Pressure (kPa)	101	101
Fluid package	Peng-Robinson	

Table 2. Aspen HYSYS input condition for alternative 2

Stream name	Water 1	Water 6
Temperature (ºC)	65	80
Pressure (kPa)	101	101
Fluid package	Peng-Robinson	

An Aspen HYSYS flow-sheet model of the process for the simulation is presented in Figure 2.
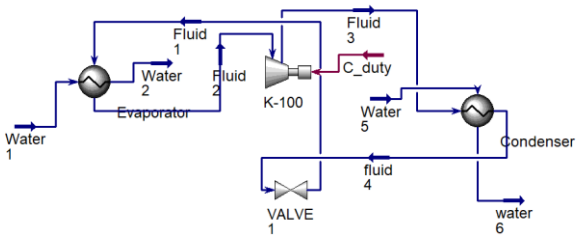
Figure 2. Model representation of the DC heat recovery process via heat pump in Aspen HYSYS

Water 1 is the cooling water from the data center and water 6 is water supplied to the district heating network. The result of the simulation was obtained in a very short time as the simulation in Aspen HYSYS is very quick and efficient. The simulation results for the two alternatives are presented in Table 3 and Table 4.

Table 3. Results of material and energy balance achieved from Aspen HYSYS for alternative 1

	Wat 1	Fluid 1	Wat 2	Fluid 2	Fluid 3	Wat 5	Wat 6	Fluid 4
T (°C)	65	51.24	55	51.24	88.84	40	70	74.25
P (kPa)	101	2000	101	2000	3280	101	101	3280
Flow (kgmole/h)	55.	4.369	55.51	4.369	4.369	20.94	20.94	4.369
Flow (kg/h)	1000	377.8	1000	377.8	377.8	377.1	377.1	377.8
Liq flow (m³/h)	1.0	0.308	1.002	0.308	0.308	0.377	0.377	0.308
Heat flow (kJ/h)	1.5e+7	2.131 e+06	1.576 e+07	2.088 e+06	2.082 e+06	5.968 e+06	5.919 e+06	2.131 e+06

Table 4. Results of material and energy balance achieved from Aspen HYSYS for alternative 2

	Water 1	Fluid 1	Water 2	Fluid 2	Fluid 3	Water 5	Water 6	Fluid 4
T (°C)	65	51.24	55	51.24	104.7	40	80	84.33
P (kPa)	101	2000	101	2000	4000	101	101	4000
Flow (kgmole/h)	55.	5.419	55.51	5.419	5.419	17.02	17.02	5.419
Mass flow (kg/h)	1000	468.6	1000	468.6	468.6	306.6	306.6	468.6
Liq flow (m³/h)	1.0	0.382	1.002	0.382	0.382	0.307	0.307	0.382
Heat flow (kJ/h)	1.5e+7	2.633 e+06	1.576 e+07	2.590 e+06	2.580 e+06	4.851 e+06	4.798 e+06	2.633 e+06

4.3 Calculation of COP for heat pump

For alternative 1
The evaporation temperature from the simulation is found 51.24°C.
Condensation temperature from the simulation is found 74.25°C.

From the simulation the amount of heat output from the condenser, Q_C = 48950 kJ/h.
From the simulation power required in the compressor, W = 5651kJ/h.

$$COP = \frac{Q_C}{W} = \frac{48950}{5651} = 8.66$$

For alternative 2
The evaporation temperature from the simulation is found 51.24°C.
Condensation temperature from the simulation is found 84.33°C.
From the simulation the amount of heat output from the condenser, Q_C = 53130 kJ/h.
From the simulation power required in the compressor, W = 9829 kJ/h.

$$COP = \frac{Q_C}{W} = \frac{53130}{9829} = 5.4$$

So, when the cooling water from the data center is 65°C and the supply water to district heating is 70°C the COP is found 8.66. On the other hand, when the cooling water from the data center is 65°C and the supply water to district heating is 80°C the COP is found 5.4. The two COP values can be compared to an average performance COP value of 6.8 from Oltmanns et al. (2020). In that case the cooling water input was 45 °C and the return temperature for the heat delivery system varied between 50 and 70 °C.

4.4 Economic calculation

For the energy cost calculation, simple assumptions are made. The price of electricity is estimated to be 0.1 EUR/kWh, and the district heat price was specified to 0.05 EUR/kWh.
So, the formula for the estimated economic potential is presented in equation 3.

$$\begin{aligned} Economic\ potential \quad\quad (3)\\ = Price \cdot Recovered\ heat\\ - (Elc.price\\ \cdot \frac{Recovered\ heat}{COP}) \end{aligned}$$

Oltmanns *et al.* (2020) have found that in 2018 the Telia data center in Helsinki, Finland supplied 200GWh/a in the nearby city of Espoo. So, taking this recovered heat value as a reference, the economy for a large data center facility can be calculated.

For alternative 1
Economic potential $= 0.05 \frac{EUR}{kWh} \cdot 200GWh - \left(0.1 \frac{EUR}{kWh} \cdot \frac{200GWh}{8.66}\right) = 7.7$ MEUR

For alternative 2

Economic potential $= 0.05 \frac{EUR}{kWh} \cdot 200 GWh - (0.1 \frac{EUR}{kWh} \cdot \frac{200 GWh}{5.4}) = 6.3$ MEUR

For the case of omitting heat pump, all the 200GW energy can be utilized to district heating network, which is worth of value 10 MEUR, as per kW district heat price is 0.05 EUR.

The investment cost is mostly dependent on the installation cost of the heat pump facility. Other costs can be negligible for the heat recovery solution. The heat pump cost is very critical to determine. From the study of Nishihata et al. (2013) the installed cost for a 2 kW pump is found to be 552.4 EUR. Thus, for a 24 MW capacity data center, the cost of the heat pump will be equal to 6.6 MEUR. Hence for a large amount of heat recovery from DC, it can be estimated that the investment of heat pump cost is high enough.

If the project is run for 10 years, the economic value for all three alternatives to be calculated. The factor for constant income is given by equation

$$Factor = \frac{1 - (1 + i)^{-n}}{i} \quad (4)$$

For n = 10 years and i = 7%, the factor is 7.02.
The economic result is to be calculated by equation 5

$$
\begin{aligned}
Economic\ result \quad &(5)\\
&= Economic\ potential\\
&\cdot Factor\\
&- Investment\ cost
\end{aligned}
$$

For alternative 1
The economic result = 7.7 ·7.02-6.6 = 47.45 MEUR
For alternative 2
The economic result = 6.3 ·7.02-6.6 = 37.63 MEUR
For the alternative without a heat pump, there will be no investment cost. So, the economic result will be 70.2 MEUR.
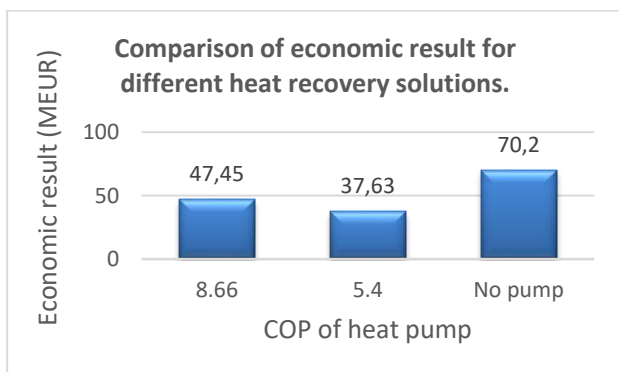
Figure 3. COP of pump versus the economic value of heat recovery solutions

The comparison of the economic results for different heat recovery scenarios is represented in Figure 3.

4.5 Evaluation of uncertainties

The three alternative solutions for heat recovery from a data center depend on certain things. The most vital thing is the temperature dependency. The required cooling temperature specified by data centers and the required temperature for the district heating network should be specified by the respective companies. In future work, it would be interesting to continue this work based on data from existing data centers. These two temperatures play a vital role in heat recovery solutions. If the temperatures change the value of the economic result will also vary. Besides the prices of electricity and district heating may vary country wise which will also affect the result. For the required calculation, they are assumed. Moreover, pipeline cost is very difficult to estimate as it depends on the climate, length of connection, and environmental condition. However, it can be assumed that the heat pump cost is relatively larger than the pipeline cost.

5 Conclusion

Simulations and economical optimization at different conditions in Aspen HYSYS were carried out. Especially three alternatives were evaluated. The first is an alternative without a heat pump in which the cooling water leaves the data center at 80 ℃ and enters the district heat network at 70 ℃. The second is an alternative with a slight temperature increase in the heat pump. The cooling water temperature from the data center is 65 ℃ and the temperature to the district heat system is 70 ℃. The third is an alternative with a higher temperature increase in the heat pump. The cooling water temperature from the data center is 65 ℃ and the temperature to the district heat system is 80 ℃. The COP (Coefficient of Performance) in a heat pump for these alternatives were calculated using the refrigerant R-22 in the simulation program Aspen HYSYS. The estimated economic potential for each alternative was calculated by estimated values on electricity cost and district heat price. In one alternative, the electricity cost was specified to 0.1 EUR/kWh, and the district heat price was specified to 0.05 EUR/kWh. For the alternatives using heat pumps, the capital cost was estimated assuming that the heat pump investment was dominating.

The COPs for the two heat pump alternatives were calculated to be 8.66 and 5.4, respectively. The economy for a large data center facility with recovered waste heat of 200 GWh/year was calculated for 10 years.

For the specified conditions, the net present value was calculated to be large and positive for all the alternatives. As expected, the most economical alternative was without a heat pump, and the most economical heat pump was the one with the highest COP. Pipeline cost is very much dependent on the length and the local conditions for which it was not possible to make a reasonable estimation.

The calculations show that there is a large potential in using waste heat from data centers for district heating.

References

A. Almoli, A. Thompson, N. Kapur, J. Summers, H. Thompson, and G. Hannah. Computational fluid dynamic investigation of liquid rack cooling in data centres, *Applied energy,* 89(1):150-155, 2012. doi:10.1016/j.apenergy.2011.02.003

ASHRAE. Thermal guidelines for data processing environments–expanded data center classes and usage guidance. *Whitepaper prepared by ASHRAE technical committee (TC),* 9, 2011.

L. Bellia, A. Capozzoli, P. Mazzei, and F. Minichiello. A comparison of HVAC systems for artwork conservation. *International Journal of Refrigeration,* 30(8): 1439-1451, 2007. doi:10.1016/j.ijrefrig.2007.03.005

T. Brunschwiler, B. Smith, E. Ruetsche, and B. Michel. Toward zero-emission data centers through direct reuse of thermal energy. *IBM Journal of Research and Development,* 53(3):1-11, 2009. doi: 10.1147/JRD.2009.5429024

A. Capozzoli, M. Chinnici, M. Perino, and G. Serale. Review on performance metrics for energy efficiency in data center: The role of thermal management. In *International Workshop on Energy Efficient Data Centers*, Springer, pages 135-151, 2014. doi: 10.1007/978-3-319-15786-3_9

A. Capozzoli and G. Primiceri. Cooling systems in data centers: state of art and emerging technologies. *Energy Procedia,* 83:484-493, 2015. doi: 10.1016/j.egypro.2015.12.168

A. Capozzoli, G. Serale, L. Liuzzo, and M. Chinnici. Thermal metrics for data centers: A critical review. *Energy Procedia,* 62:391-400, 2014. doi:10.1016/j.egypro.2014.12.401

Y. Q. Chi *et al.* Case study of a data centre using enclosed, immersed, direct liquid-cooled servers. In *Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM)*, 2014: IEEE, pages 164-173, 2014. doi:10.1109/SEMI-THERM.2014.6892234

K. Dunlap and N. Rasmussen. Choosing between room, row, and rack-based cooling for data centers. *APC White Paper,* 130, 2012.

K. Ebrahimi, G. F. Jones, and A. S. Fleischer. A review of data center cooling technology, operating conditions and the corresponding low-grade waste heat recovery opportunities. *Renewable and Sustainable Energy Reviews,* 31:622-638, 2014. doi: 10.1016/j.rser.2013.12.007

H. Geng. *Data center handbook.* John Wiley & Sons, 2014.

M. Iyengar *et al.* Server liquid cooling with chiller-less data center design to enable significant energy savings. In *28th annual IEEE semiconductor thermal measurement and management symposium (SEMI-THERM)*, IEEE, pages 212-223, 2012. doi:10.1109/STHERM.2012.6188851

M. Iyengar *et al.* Extreme energy efficiency using water cooled servers inside a chiller-less data center. In *13th InterSociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems*, IEEE, pages 137-149, 2012. doi: 10.1109/ITHERM.2012.6231424

J. Niemann, K. Brown, and V. Avelar. Impact of hot and cold aisle containment on data center temperature and efficiency. *Schneider Electric Data Center Science Center, White Paper,* 135, pages 1-14, 2011.

H. Nishihata, R. Kempener, G. Simbolotti, and G. Tosato, Heat Pumps—Technology Brief. *International Energy Agency (IEA)—Energy Technology Systems Analysis Programme (ETSAP)/International Renewable Energy Agency (IRENA): Paris, France,* page 24, 2013.

M. Ohadi, S. Dessiatoun, K. Choo, M. Pecht, and J. V. Lawler. A comparison analysis of air, liquid, and two-phase cooling of data centers. In *28th Annual IEEE Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM)*, IEEE, pages 58-63, 2012. doi: 10.1109/STHERM.2012.6188826

J. Oltmanns, D. Sauerwein, F. Dammel, P. Stephan, and C. Kuhn. Potential for waste heat utilization of hot-water-cooled data centers: A case study. *Energy Science & Engineering,* 2020. doi:10.1002/ese3.633

E. Oró, R. Allepuz, I. Martorell, and J. Salom. Design and economic analysis of liquid cooled data centres for waste heat recovery: A case study for an indoor swimming pool. *Sustainable cities and society,* 36:185-203, 2018. doi:10.1016/j.scs.2017.10.012

C. D. Patel. A vision of energy aware computing from chips to data centers. In *The international symposium on micro-mechanical engineering*, 2003. doi:ISMME2003-K15

R. R. Schmidt, E. E. Cruz, and M. Iyengar. Challenges of data center thermal management. *IBM Journal of Research and Development,* 49(4.5):709-723, 2005. doi: 10.1147/rd.494.0709

C. S. Sharma, S. Zimmermann, M. K. Tiwari, B. Michel, and D. Poulikakos. Optimal thermal operation of liquid-cooled electronic chips. *International journal of heat and mass transfer,* 55(7-8):1957-1969, 2012. doi: 10.1016/j.ijheatmasstransfer.2011.11.052

Simple, Science, and Technology. (2011, 15.05.2020). *Refrigeration Cycle* [Online]. Available: http://simple-engineering.blogspot.com/2011/01/refriger%20nvbfc%20dxz%C3%A5ation-cycle.html.

R. Smith, *Chemical process: design and integration.* John Wiley & Sons, 2005.

S. Zimmermann, I. Meijer, M. K. Tiwari, S. Paredes, B. Michel, and D. Poulikakos. Aquasar: A hot water cooled data center with direct energy reuse. *Energy,* 43(1):237-245, 2012. doi: 10.1016/j.energy.2012.04.037

L. E. Øi and I. Y. Tirados. Heat Pump Efficiencies simulated in Aspen HYSYS and Aspen Plus. 2015. doi: 10.3384/ecp15119141

Study of an Industrial Electrode Dryer of a Lithium-Ion Battery Manufacturing Plant: Dynamic Modeling

Emil Oppegård[1]   Asanthi Jinasena[2]   Anders Hammer Strømman[2]   Jon Are Suul[1,3]   Odne Stokke Burheim[2]

[1]Department of Engineering Cybernetics, Norwegian University of Science and Technology, Norway,
`emilop@stud.ntnu.no`, `jon.are.suul@ntnu.no`
[2]Department of Energy and Process Engineering, Norwegian University of Science and Technology, Norway,
`{asanthi.jinasena,anders.hammer.stromman,odne.s.burheim}@ntnu.no`
[3]SINTEF Energy Research, Norway

Abstract

A dynamic model for lithium-ion battery (LIB) electrode manufacturing and drying is developed in this paper. The model is intended for analysis of different drying technologies, energy requirement calculations, and optimization and control of the drying process. The model shows that the infrared drying is faster than the convective drying when the heat source temperature is the same. The energy required to evaporate the solvent can be reduced by gradually changing the hot air temperature. Drying is the most energy-intensive process in cell manufacturing, and the cell manufacturing process is the biggest contributor to greenhouse gas emissions in the LIB industry. Therefore, the presented model is useful for accurate estimation of the environmental impact as well as for identifying the appropriate measures to reduce energy requirements in the rapidly growing LIB industry.

Keywords: lithium-ion battery, electric vehicle, electrode drying, convection, infrared radiation, sustainable energy, model, control

1 Introduction

Efforts to decarbonize the transport sector are currently leading to a shift from fossil fuelled vehicles to electric mobility. Although this development is still at an early stage, the global market for electric mobility is expected to grow rapidly in the future. Currently, the largest uptakes of electric vehicles (EVs) are appearing in China and Europe. Within Europe, the Nordic countries have a leading position. Norway has a high share of EVs, where 40% of all new car sales in 2019 were battery electric vehicles (BEVs) (International Energy Agency, 2019).

The demand for LIB is expected to increase rapidly with the emerging technology for electric mobility. Ensuring that this development fulfills its potential for climate change mitigation is important. In that regard, life cycle assessment studies provide good insights. They have shown that the carbon footprint from LIB production may contribute to a significant part of the overall greenhouse gas emissions of BEVs (Kurland, 2020; Ellingsen et al.,

2016). Especially, energy requirements in the cell manufacturing process are important drivers for these emissions (Kurland, 2020).

A thorough understanding of the energy requirements of the cell manufacturing process is crucial for an accurate assessment of the current practices, the identification of appropriate measures to reduce energy requirements, and the maximization of the potential for mitigating climate change. Thus, there is a need for research driving towards sustainable, low cost, efficient, and high energy density battery manufacturing practices. Consequently, the use of sustainable materials, the adaption of less energy-intensive technologies, and energy benchmarking of process control practices are becoming important (Susarla et al., 2018). An in-depth understanding of the underlying principles of battery manufacturing would give valuable insight into the fulfillment of these requirements.

The LIB consists of electrodes (cathode and anode), separator, electrolyte, and a casing, for more details see Burheim (2017). Both the cathode and anode are made of active materials, carbon black additives, and a polymeric binder. Usually, the active material for the cathode is lithium metal oxides such as $LiMn_2O_4$, $Li_aNi_xMn_yCo_zO_2$. For either cathode or anode, carbon black conductive additives are acetylene black, or graphite while the polymeric binder is polyvinylidene difluoride (Kwade et al., 2018). These materials are dry-mixed together and combined with a solvent to make a slurry. The solvent is either N-Methyl-2-pyrrolidone (NMP), or water. Once the slurry is made, it is coated onto a thin aluminium (cathode slurry) or copper (anode slurry) foil, typically of 25-35 μm in thickness (Burheim, 2017). These coated thin films are then dried to remove most of the solvent. The dried films are then pressed to gain the predefined porosity and thickness, which is known as calendaring. Then inside a dry room, the electrodes are cut, stacked, and assembled into the cell, with the rest of the parts. This cell is then activated by cell conditioning processes.

Out of these, the most energy-intensive step is the elec-

trode drying process (Yuan et al., 2017; Dai et al., 2019), which also highly influences the cell performance by the drying rate and method (Ahmed et al., 2016). A physics-based model of this process would not only give a good insight into dynamic drying behavior and energy optimization, but also the potential of using new materials, different compositions, and different drying techniques. Although modeling of thin-layer drying is a well-researched topic in food, paper, paint, and pharmaceutical industries, there is limited available literature dedicated to LIB manufacturing industry. The few existing models (Ahmed et al., 2016; Susarla et al., 2018) are insufficient to fulfill the above requirements. Therefore, we develop a dynamic mathematical model of the electrode drying process. The presented model is intended for analysis of different drying technologies, including calculation of energy requirement as well as optimization and control of the process. In addition, a sensitivity analysis of process parameters is presented to evaluate the model performance.

The paper is organized as follows. The system description including the model derivation, the used numerical method and its' stability are presented in Section 2. This is followed by detailed results and discussions being presented in Section 3. Finally the conclusions that are drawn from the study are summarized in Section 4.

2 Model Development

Drying processes are inherently complex due to simultaneous mass and heat transfer. The shrinkage of the thin film further complicates the process due to the formation of a porous structure as the solvent evaporates. The drying process has two stages; in the first stage evaporation is only a surface phenomenon and the thin film shrinks. In the second stage, the film does not shrink, evaporation is capillary-driven, hence no longer limited to the surface (Susarla et al., 2018). Further, during the first phase, the active material particles and binder are dispersed in the solvent. However, in the second phase, the particles are drawn close to each other and the voids created by solvent vaporization get replaced by air. Since the evaporation is governed by different phenomena, the two distinct stages require different mathematical descriptions. For this study, the developed model will emphasize only on the first stage.

2.1 Model Derivation

To derive a simple, yet accurate model that describes the drying process which later can be used in optimization and control, the problem is simplified by well-established assumptions and simplifications. Prior to the drying process, the slurry is assumed to be uniformly mixed and coated onto the current collectors. Thus, the temperature and slurry distribution in the x-direction (width of the film) and in the y-direction (length of the film) are uniform within the control volume. The electrode is assumed to be moving in the y-direction with a constant velocity. This leaves the control volume to only consider fluxes in the z-direction (height of the film), which results in a one-dimensional model. For further simplification, the thickness of the film is assumed to be so small that the temperature development in the z-direction is uniform.

The fluxes in the z-direction due to diffusion can be expressed by the one-dimensional diffusion equation:

$$\frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial z^2} \tag{1}$$

where, C represents the weight fraction of the solvent, t is the time and D is the diffusion coefficient. The diffusion coefficient is highly dependent on the temperature and concentration. Mesbah et al. (2014) expresses the diffusion coefficient as a function of polymer volume fraction and temperature as follows,

$$D = D_0 \cdot \left(\frac{1 - \phi_p}{1 + \phi_p} \right)^{\gamma} \cdot \exp\left(\frac{-E}{RT} \right). \tag{2}$$

where D_0 refers to the mutual diffusion coefficient and varies with different compositions, ϕ is the polymer volume fraction, E is the activation energy, R is the universal gas constant, γ is an empirical value and T is the film temperature. The liner is assumed to be impenetrable, which is equivalent to imposing a zero flux boundary condition at the bottom of the film:

$$z = 0: \quad \frac{\partial C}{\partial z} = 0 \tag{3}$$

For the boundary at the top of the film, the boundary condition is governed by the evaporation of the solvent. The shrinkage of the film also has to be taken into consideration. Mesbah et al. (2014) expressed the boundary condition as:

$$z = h(t): \quad D \frac{\partial C}{\partial z} + C \frac{dh}{dt} = -\frac{k_m M_s}{\rho_s R} \cdot \left(\frac{P_s}{T} - \frac{P_a}{T_a} \right) \tag{4}$$

where, h is the varying thickness of the film, k_m is the mass transfer coefficient, M_s is the molecular weight of the solvent, ρ_s is the density of the solvent, P_s is the partial pressure of the solvent, P_a is the vapor pressure in the air and T_a is the temperature of the heated air. The vapor pressure and temperature of the heated air is considered to be constant throughout the drying process. The vapor pressure is approximated using the Antoine equation:

$$P_a = \varphi \cdot 133.332 \cdot 10^{\left(7.54826 - \frac{1979.68}{222.2 + (T_a - 273.15)} \right)} \tag{5}$$

where φ represents the relative humidity of the heated air. In contrast to the vapor pressure, P_s varies with both solvent concentration and temperature. Khansary (2016) expresses the solvent vapor pressure using Flory Huggin's theory for polymeric solutions as follows,

$$P_s = P_0 \cdot \exp\left(\phi_p^2 \mathscr{X} + \ln(\phi_s) + \left(1 - \frac{V_s}{V_p} \right) \phi_p \right), \tag{6}$$

where, P_0 is the vapor pressure of the pure solvent, $\mathscr{X}$ is the Flory Huggins interaction parameter, ϕ_s represents the volume fraction of the solvent, and V_s, V_p are the volumes of solvent and solid, respectively.

An expression for the thickness of the film is obtained by applying mass balance to the system. Since the thickness is solely dependent on the evaporation of the solvent, it can be described by the following formula:

$$\frac{dh}{dt} = -\frac{k_m M_s}{\rho_f R} \cdot \left(\frac{P_s}{T} - \frac{P_a}{T_a} \right), \qquad (7)$$

where, ρ_f represents the density of the film. The densities of the combined solids and solvent are assumed to be constant throughout the drying process. However, the density of the film is expressed as a linear function of the individual densities and solvent mass fraction:

$$\rho_f = \rho_s \cdot C + (1 - C) \cdot \rho_p, \qquad (8)$$

where, ρ_p is the bulk density of the solids.

Energy balance is applied to the system with the stated assumptions. In addition, contributions from surface tension and shear stress gradients are assumed to be so small that they can be neglected without affecting the accuracy significantly. The energy balance for the film is as follows:

$$h\frac{dT}{dt} + T\frac{dh}{dt} = -\frac{k_m M_s \lambda}{\rho_f C_p R} \cdot \left(\frac{P_s}{T} - \frac{P_a}{T_a} \right) + \frac{\dot{q}_{in}}{\rho_f C_p} \qquad (9)$$

where, λ is the latent heat of vaporization of the solvent and can be approximated by a second degree polynomial as follows:

$$\lambda = 6.991 \cdot T^2 - 6193 \cdot T + 1.848 \cdot 10^6 : \text{NMP} \qquad (10)$$

$$\lambda = -3.345 \cdot T^2 - 259.3 \cdot T + 2.817 \cdot 10^6 : \text{water} \qquad (11)$$

C_p is the specific heat of the film. The variation of C_p within the temperature range is insignificant. Therefore, it is set to a constant for simplicity. $\dot{q}_{in}$ represents the heat transferred into the film. The heat transportation is either done by forced convection, radiation or a combination of both depending on the configuration of the drying process. Heat transportation done by convection is given by:

$$\dot{q}_{in} = k_c \cdot (T_a - T) \qquad (12)$$

where, k_c is the heat transfer coefficient. Susarla et al. (2018) determines the heat transfer coefficient based on the air velocity as follows:

$$k_c = 0.037 \cdot v^{0.8} \cdot \left(\frac{\mu}{\rho_a} \right)^{-0.8} \cdot \text{Pr}^{1/3} \cdot L^{-0.2}, \qquad (13)$$

where, v is the air velocity, μ is the viscosity of the air, ρ_a is the density of the air, Pr is the Prandtl number and L is the characteristic length. Only the convective heat transfer is considered during convective drying, and the

temperature of the hot air is assumed to be unchanged by the effects of the solvent evaporation.

However, for infrared radiative drying, the radiative heat transfer from the source to the film, and the convective heat transfer from the source to the air stream above the film are considered. This is because the airflow rate is considered to be at ambient conditions, hence the temperature of the air changes with heat transfer from the source, and the solvent evaporation. Heat transportation from the source to the film due to radiation can be written as follows,

$$\dot{q}_{in} = \sigma \varepsilon \cdot \left(T_r^4 - T^4 \right), \qquad (14)$$

where, T_r is the heat source temperature, σ is the Stefan Boltzman constant and ε is the emissivity of the surface of the film. The energy required for solvent evaporation throughout the process is given by:

$$Q = \int_0^{t_f} \dot{m}_{evap} \cdot \lambda \, dt \qquad (15)$$

where, t_f is the final time of the drying process and $\dot{m}_{evap}$ is the evaporation rate. The total energy requirement for drying would include the energy needed for the heat source, in addition to Q.

In addition, the air temperature during the infrared drying can be calculated by the energy balance to a unit volume of air stream as follows,

$$\frac{dT_a}{dt} = \frac{k_c(T_r + T - 2T_a) + \dot{m}_{evap}C_{ps}(T - T_a)}{\rho_a C_{pa}}, \qquad (16)$$

where C_{ps} and C_{pa} are specific heat capacities of the solvent and air, respectively.

To complete the model, the initial conditions at $t(0)$ are given by:

$$C(0, z) = C_0, \quad T(0) = T_0, \quad h(0) = h_0 \qquad (17)$$

2.2 Dimensionless Model

To simplify the numerical approach by immobilizing the otherwise moving boundary conditions at the surface, dimensionless variables are introduced. The variables can be expressed by the following dimensionless variables.

$$\tilde{t} := \frac{t}{t_0} \qquad \tilde{z} := \frac{z}{h} \qquad \tilde{h} := \frac{h}{h_0} \qquad (18)$$

$$\tilde{T} := \frac{T}{T_0} \qquad \tilde{C} := \frac{C}{C_0} \qquad \tilde{D} := \frac{D}{D_0} \qquad (19)$$

where, $\tilde{z}$, $\tilde{h}$, $\tilde{T}$, $\tilde{C}$, $\tilde{D}$ represent the dimensionless variables which ranges from 0 to 1, while $\tilde{t} = t$, by setting t_0 to unity for simplicity. Substituting equations 18 and 19 into the continuous model yields the following dimensionless model:

$$\frac{d\tilde{h}}{d\tilde{t}} = -\frac{k_m M_s t_0}{\rho_f R h_0} \cdot \left(\frac{P_s}{\tilde{T}T_0} - \frac{P_a}{T_a} \right), \qquad (20)$$

$$\frac{\mathrm{d}\tilde{T}}{\mathrm{d}\tilde{t}} = -\frac{k_\mathrm{m} M_\mathrm{s} \lambda t_0}{\rho_\mathrm{f} C_\mathrm{p} R \tilde{h} h_0} \cdot \left(\frac{P_\mathrm{s}}{\tilde{T} T_0^2} - \frac{P_\mathrm{a}}{T_\mathrm{a} T_0} \right) - \frac{\tilde{T}}{\tilde{h}} \frac{\mathrm{d}\tilde{h}}{\mathrm{d}\tilde{t}}$$
$$+ \frac{t_0}{\rho_\mathrm{f} C_\mathrm{p} \tilde{h} h_0 T_0} \dot{q}_\mathrm{in}, \tag{21}$$

$$\frac{\partial \tilde{C}}{\partial \tilde{t}} = \frac{\tilde{D} D_0 t_0}{(\tilde{h} h_0)^2} \frac{\partial^2 \tilde{C}}{\partial \tilde{z}^2}, \tag{22}$$

with the boundary conditions:

$$\tilde{z} = 0: \quad \frac{\partial \tilde{C}}{\partial \tilde{z}} = 0 \tag{23}$$

$$\tilde{z} = 1: \quad \frac{\tilde{D} D_0 C_0}{\tilde{h} h_0} \frac{\partial \tilde{C}}{\partial \tilde{z}} + \frac{\tilde{C} C_0 h_0}{t_0} \frac{\mathrm{d}\tilde{h}}{\mathrm{d}\tilde{t}} =$$
$$- \frac{k_\mathrm{m} M_\mathrm{s}}{\rho_\mathrm{s} R} \left(\frac{P_\mathrm{s}}{\tilde{T} T_0} - \frac{P_\mathrm{a}}{T_\mathrm{a}} \right) \tag{24}$$

and the initial conditions at $t(0)$:

$$\tilde{C}(0, \tilde{z}) = 1, \quad \tilde{T}(0) = 1, \quad \tilde{L}(0) = 1. \tag{25}$$

2.3 Numerical Approach and Stability

The system is solved in MATLAB® using the finite difference method (FDM). FDM approximates the derivatives using the finite part of the Taylor Series expansion (Tannehill et al., 1997). Before applying the FDM, the domain of the system is discretized into a finite grid of nodes, with the nodes representing the system properties at a finite point in both space and time. The forward-time-central-space (FTCS) differential scheme is used in approximating the derivatives as follows (Tannehill et al., 1997):

$$\frac{\mathrm{d}f}{\mathrm{d}\tilde{t}} = \frac{f^{i+\Delta \tilde{t}} - f^i}{\Delta \tilde{t}} \tag{26}$$

$$\frac{\partial^2 f}{\partial \tilde{z}^2} = \frac{f_{j+\Delta \tilde{z}} - 2f_j + f_{j-\Delta \tilde{z}}}{\Delta \tilde{z}^2} \tag{27}$$

Applying the FDM with FTCS-scheme yields the following discrete equations:

$$\tilde{h}^{i+\Delta \tilde{t}} = -t_0 \Delta \tilde{t} \frac{k_\mathrm{m} M_\mathrm{s}}{\rho_\mathrm{f} R h_0} \cdot \left(\frac{P_\mathrm{s}}{\tilde{T}^i T_0} - \frac{P_\mathrm{a}}{T_\mathrm{a}} \right) + \tilde{h}^i \tag{28}$$

$$\tilde{T}^{i+\Delta \tilde{t}} = -t_0 \Delta \tilde{t} \frac{k_\mathrm{m} M_\mathrm{s} \lambda}{\rho_\mathrm{f} C_P R \tilde{h}^i h_0} \cdot \left(\frac{P_\mathrm{s}}{\tilde{T}^i T_0^2} - \frac{P_\mathrm{a}}{T_\mathrm{a} T_0} \right) +$$
$$\frac{t_0 \Delta \tilde{t}}{\rho_\mathrm{f} C_P \tilde{h}^i h_0 T_0} \cdot \dot{q}_\mathrm{in} - \frac{\tilde{T}^i \tilde{h}^{i+\Delta \tilde{t}}}{\tilde{h}^i} + 2\tilde{T}^i \tag{29}$$

$$\tilde{C}_j^{i+\Delta \tilde{t}} = \frac{\tilde{D}_j^i D_0 t_0 \Delta \tilde{t}}{(\tilde{h}^i h_0)^2} \cdot \frac{\tilde{C}_{j+\Delta \tilde{z}}^i - 2\tilde{C}_j^i + \tilde{C}_{j-'\Delta \tilde{z}}^i}{\Delta \tilde{z}^2} + \tilde{C}_j^i \tag{30}$$

The upper boundary is found using a ghost cell.

There is a trade-off between the accuracy and the stability of the numerical method. The dimensionless diffusion equation is the limiting term for the stability of this system. Thus, the stability criterion for the system can be expressed by the general stability criterion for FDM (Tannehill et al., 1997) on the diffusion equation as follows:

$$\frac{\Delta \tilde{t}}{\Delta \tilde{z}^2} \leq \frac{1}{2 \max \left(\frac{\tilde{D} D_0 t_0}{(\tilde{h} h_0)^2} \right)} \tag{31}$$

Therefore, the selection of the resolution is done to get a stable and accurate enough solution.

3 Results and Discussion

The simulations for the composition of a 110 Wh (around 30Ah), NMC811-G battery is carried out for selected process parameters. The film thickness, film temperature, solvent fraction, and the drying rate with time is plotted together in Figure 1. A hot air stream and a radiating heat source of the same temperature are used to test both the convective drying and infrared radiation drying. Only cathode drying with NMP as solvent is presented here for brevity, except otherwise stated.

As shown in Figure 1, The temperature of the film rapidly reaches the hot air temperature during infrared drying, and within 2 minutes for the convective drying, while the solvent evaporation is continued for about 90 minutes for both drying processes. The film thickness reached a constant value at the end of the drying period, where approximately 90% and 99% of the solvent is evaporated during convective and radiative drying, respectively. This is because the radiative drying gives a higher drying rate and shorter drying time than convective drying. However, it requires more energy for solvent removal (Q) than the convective drying.

The general behavior of solvent removal and film thickness reduction as shown in Mesbah et al. (2014), can be seen here as well. Further, the film temperature transients and the drying rate are of a similar pattern to the model results given in Susarla et al. (2018). Moreover, Susarla et al. (2018) have used the capillary pressure together with the vapor pressure in the model, which gives high drying rates and low drying times. However, Ahmed et al. (2016) suggest that the cathode drying can be limited by the allowable NMP concentration in the dryer outlet air (1000-2000 ppm), which means that the cathode drying rate has an upper limit controlled by the dryer inlet air conditions and the initial film conditions. Further, Rollag et al. (2019) observed that increased electrode thickness and elevated drying temperatures cause crack formation in aqueous cathodes. These cracks were formed during the second stage of drying, which is driven by capillary forces. Therefore, improvements to the model to address these issues need to be thoroughly investigated in the future.

3.1 Effect of Initial and Process Parameters

The initial parameters such as the initial solvent mass fraction C_0 and the initial film thickness h_0 are defined by the type of battery that is being produced. However, the initial film temperature T_0, and the reference diffusion co-
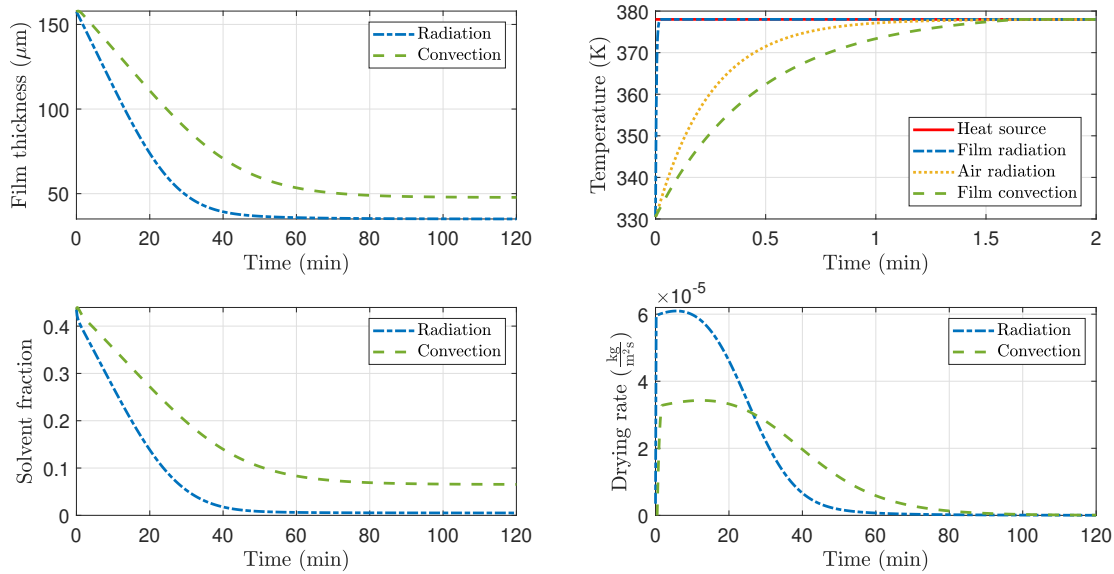
Figure 1. Film properties of the cathode during the convection and radiation drying processes.

efficient D_0 mainly depend on the indoor environmental conditions. The process parameters, which are the properties of the heat source (hot air or infrared heater) can be changed selectively. For convective drying, these parameters are the relative humidity φ, temperature T_a, and the velocity v of the hot air stream. For radiation, the temperature T_s and emissivity ε of the heater are the parameters that can be changed.

To evaluate the effect of these parameters on the model outputs (convective drying model), a sensitivity analysis is conducted for all the parameters. The full range of input parameter values that is suitable for the selected battery type is considered. The changes in the model output values that happen due to the changes in these input parameter values are calculated, where only one parameter changes at a time. The sensitivities are determined as a percentage variation of the output for a percentage change of each initial or input parameter. Energy required for solvent evaporation (Q calculated by Equation 15) and the drying rate are selected as outputs. The spider plot for energy requirement is shown in Figure 2.

The most influential parameter on the energy consumption is the hot air temperature (red dash dotted line) as it rapidly changes the heat transfer. Having hotter air lowers the energy consumption. The energy requirement is highly sensitive to the initial thickness of the film upon entering the dryer (purple dash dotted line) along with the initial solvent concentration of the film (blue dash dotted line). The relative humidity (water) of the hot air is also affecting the heat transfer rate due to the direct correlation of humidity with the vapor pressure. Here, in this model, the relative humidity of hot air is considered to be less than 0.5, since it is bounded by the maximum allowable limit of NMP in the outlet air of the dryer (Ahmed et al.,

2016). The other parameters have a low impact on energy consumption.

Further, the effect of the parameters on the rate of evaporation per area is calculated. All the parameters except the initial film temperature are plotted in Figure 3 since the initial film temperature has an insignificant impact on the model outputs. Change of the initial solvent concentration, the initial thickness, and the reference diffusion coefficient do not change the maximum drying rate that can be achieved. The initial solvent concentration also has little effect on the drying time. However, the initial thickness has a significant effect on the drying time, where thicker electrode coatings take a longer time to dry. Further, the lowest reference diffusion coefficient tends to have a considerably slower drying rate at the end of drying, which leads to longer drying time.

On the other hand, change of the hot air temperature, relative humidity, and air velocity change the maximum drying rate that can be achieved as well as the drying time. The direct effect of air temperature on the drying rate can
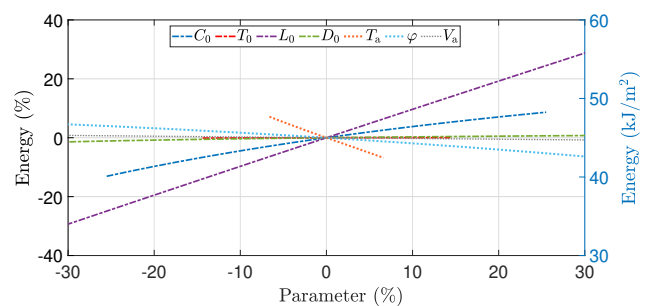
Figure 2. Effect of different initial and input parameters on the energy requirement.

also be seen here. The higher the hot air temperature, the higher the heat transfer to the electrode, thus higher the drying rate and shorter the drying time. On the contrary, when the relative humidity is increased, the evaporation of solvent to the air reduces, hence the drying rate is decreased which results in longer drying time. Energy demand can be both reduced by increasing the humidity and increasing the hot air temperature. However, since fast drying is usually preferable in the industry, low humidity is needed. The increase of air velocity tends to increase the heat transfer similar to the behaviour of air temperature, thus increases the drying rate and reduces the drying time.

Since the air properties (T_a, φ, v) are controllable parameters, these can be used to obtain faster drying. However, these parameters would also directly contribute to the energy requirements. Therefore, optimum control of the air properties is needed for the drying process to be both efficient and less energy-intensive.

3.2 Reduction of Energy Consumption

Generally in convective drying, the use of high-temperature hot air would provide a high drying rate and low drying time, which would indicate low energy requirement for the solvent evaporation. However, a large amount of hot air is needed for this, and the heat requirement for producing this high-temperature hot air is significantly higher compared to the evaporation energy (Ahmed et al., 2016). Therefore, a multi-stage drying process is tested to utilize high-temperature hot air for a shorter time, such that the heating load would be reduced.

The film properties for a three-stage drying process where the hot air temperature is gradually increased in 2-steps are shown in Figure 4. The total drying time is significantly reduced compared to the previous results, and 99% of the solvent is removed. The energy required for the solvent evaporation is lower than the single-stage use of low-temperature hot air (see Figure 5). However, the evaporation energy for the 3-stage process is higher than for the single-stage use of high-temperature air.

The energy consumption for air heating is directly proportional to the inlet and outlet temperature difference of the air heater. Assuming a constant ambient air temperature and no re-circulation of hot air, this energy is calculated for different hot air temperatures. For single-stage drying with hot air temperatures of 353 K, 368 K, and 388 K, the energy needed for air heating is 392 MJ/m^3, 326 MJ/m^3, and 276 MJ/m^3, respectively. The 3-stage drying requires 198 MJ/m^3, which is significantly less than the use of single-stage drying.

3.3 Comparison with Reported Literature Values

Considering the full range of all the parameters used in the single-stage drying, the distribution of energy requirement for convective air heating is calculated. Here, the drying of both the cathode and anode is added together. The results are compared with the reported values of energy for drying of industrial LIB plants of different capacities as shown in Figure 6.

The model values are closer to the values reported by Pettinger and Dong (2017) and Schünemann (2015), than the other values. However, the distribution of the model results has a high variation which resulted from the uncertainty in the used parameters. The range of the reported literature values is also high, due to various factors such
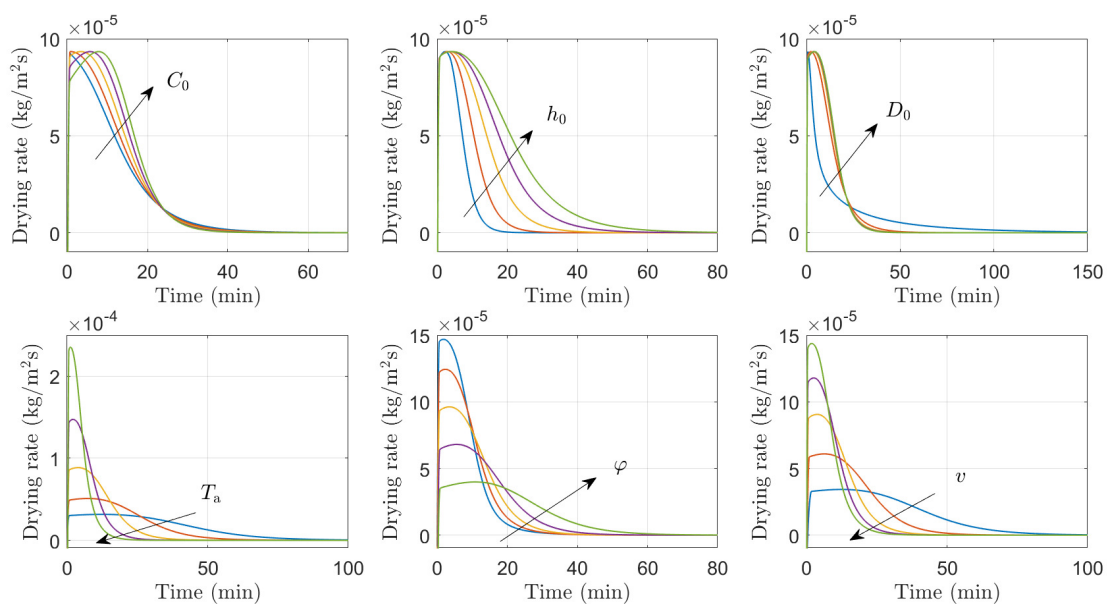
Figure 3. Effect of different initial and input parameters on the drying rate. C_0: initial solvent mass fraction, h_0: initial film thickness, D_0: reference diffusion coefficient, T_a: hot air temperature, φ: relative humidity of hot air, v: velocity of hot air
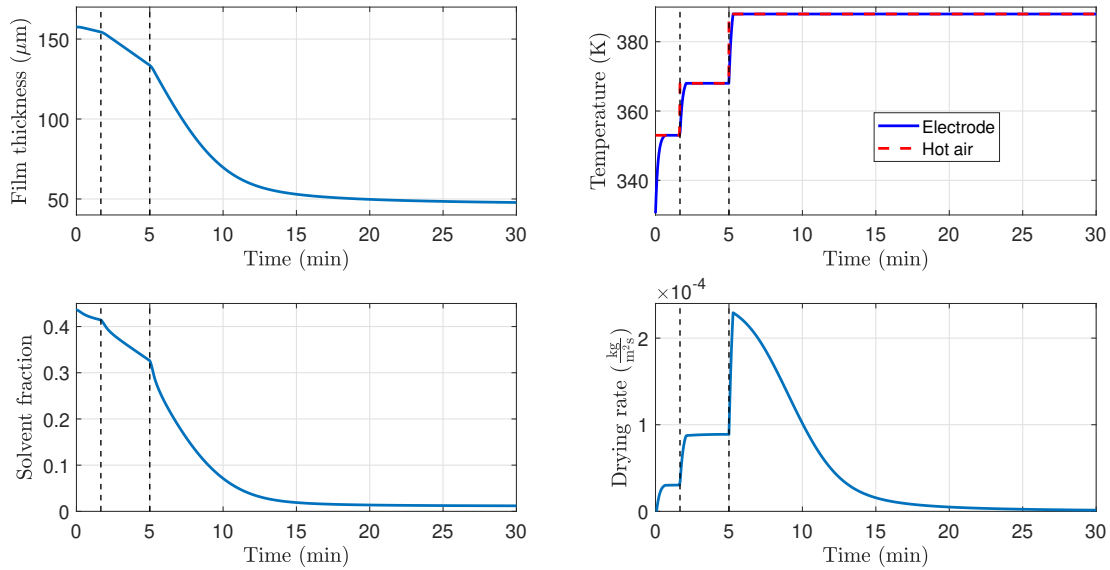
Figure 4. Film properties of cathode with the change of hot air temperature T_a.

as the annual production capacity of the plant, different battery types, and various specific energies. Further, some of the reported values are for the combined process of both calendaring and drying, as well as 2 step drying processes where the remaining solvent is vacuum dried at a later stage after calendaring. The model results provide a benchmark, however, the model does not include energy for the NMP recovery unit. No details are given regarding the literature values, whether the energy values include a NMP recovery unit or not. Ahmed et al. (2016) shows in their model, that the energy loads for solvent recovery (1969 kW) are significant compared to the energy demand for air heating (3752 kW) and solvent evaporation (130 kW).

Although the model needs improvements with the inclusion of porosity, and capillary-driven drying, the model has the capability of representing important electrode drying behaviours. Further, the model shows a good potential of being used in energy and process optimization, and process control design. Moreover, the model can be used to

compare the energy demand or drying efficiency of different drying technologies, battery chemistries, and battery types. This would further provide the opportunity to modify the model and use it with life cycle assessment models to provide updated information on electrode drying. Thus, different process modifications can be tested and assessed by the energy requirement as well as the greenhouse gas emissions and cost. It would also contribute to benchmarking the energy demand for the LIB industry.

4 Conclusions

A simple, dynamic mathematical model for electrode drying is developed for the analysis of different drying technologies and energy requirement calculations. The input uncertainty through different parameters is decomposed to determine the most influential parameters on the model output. The heat source temperature, initial thickness, and solvent concentration of the electrode are the most sen-
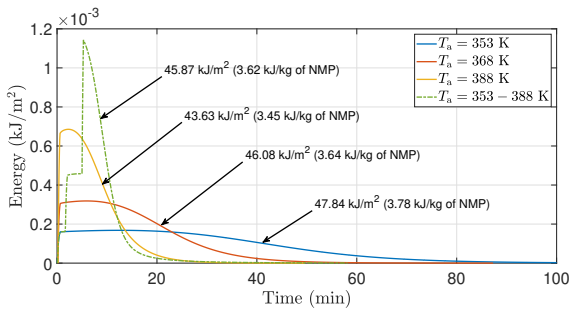
Figure 5. Energy required for the solvent evaporation at various hot air temperatures T_a. The total energy (area of each curve) for each temperature is also stated.
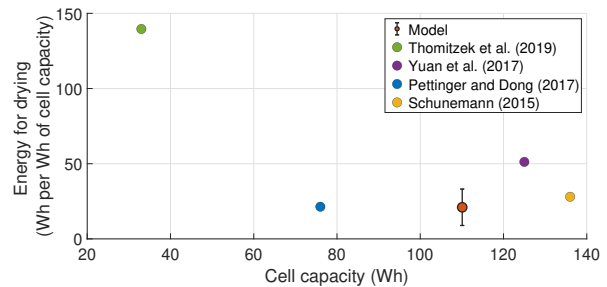
Figure 6. Comparison of energy consumption from drying model with reported studies of Schünemann (2015); Pettinger and Dong (2017); Yuan et al. (2017); Thomitzek et al. (2019). The model value is given with $\pm 2\sigma$ variation.

sitive parameters to both the drying rate and the energy consumption. Although the model is limited by the lack of porosity and capillary pressure, the results indicate the model sufficiently illustrates the general traits of an industrial electrode dryer. The results are compared with reported literature values.

Further, the use of multi-stage hot air distribution is shown to be less energy-intensive than the use of high temperature, single-stage hot air drying. The model shows a good potential of being used in energy optimization, and control design for drying processes. A further modified model can be used with life cycle assessment models. This could provide information on drying that changes with various drying technologies, energy recovery methods, and various LIB types in terms of energy requirement as well as the greenhouse gas emissions and cost.

Acknowledgments

References

Shabbir Ahmed, Paul A. Nelson, Kevin G. Gallagher, and Dennis W. Dees. Energy impact of cathode drying and solvent recovery during lithium-ion battery manufacturing. *Journal of Power Sources*, 322:169–178, 2016. doi:10.1016/j.jpowsour.2016.04.102.

Odne Stokke Burheim. *Engineering Energy Storage*. Academic press, 2017.

Qiang Dai, Jarod C. Kelly, Linda Gaines, and Michael Wang. Life cycle analysis of lithium-ion batteries for automotive applications. *Batteries*, 5(48), 2019. doi:10.3390/batteries5020048.

Linda Ager-Wick Ellingsen, Bhawna Singh, and Anders Hammer Strømman. The size and range effect: Life-cycle greenhouse gas emissions of electric vehicles. *Environmental Research Letters*, 11:1–9, 2016.

International Energy Agency. Global EV Outlook 2019. Technical report, International Energy Agency, 2019. URL www.iea.org/publications/reports/globalevoutlook2019/.

Milad Asgarpour Khansary. Vapor pressure and flory-huggins interaction parameters in binary polymeric solutions. *Korean Journal of Chemical Engineering*, 33(4):1402–1407, 2016. doi:10.1007/s11814-015-0277-6.

Simon Davidsson Kurland. Energy use for GWh-scale lithium-ion battery production. *Environmental Research Communications*, 2(1):012001, 2020. doi:10.1088/2515-7620/ab5e1e.

Arno Kwade, Wolfgang Haselrieder, Ruben Leithoff, Armin Modlinger, Franz Dietrich, and Klaus Droeder. Current status and challenges for automotive battery production technologies. *Nature Energy*, 3(4):290–300, 2018. doi:10.1038/s41560-018-0130-3.

Ali Mesbah, Ashlee N. Ford Versypt, Xiaoxiang Zhu, and Richard D. Braatz. Nonlinear model-based control of thin-film drying for continuous pharmaceutical manufacturing. *Industrial and Engineering Chemistry Research*, 53(18):7447–7460, 2014. doi:10.1021/ie402837c.

Karl Heinz Pettinger and Winny Dong. When does the operation of a battery become environmentally positive? *Journal of the Electrochemical Society*, 164(1):A6274–A6277, 2017. doi:10.1149/2.0401701jes.

Kelsey Rollag, Daniel Juarez-Robles, Zhijia Du, David L. Wood, and Partha P. Mukherjee. Drying Temperature and Capillarity-Driven Crack Formation in Aqueous Processing of Li-Ion Battery Electrodes. *ACS Applied Energy Materials*, 2(6):4464–4476, 2019. doi:10.1021/acsaem.9b00704.

Jan-Hinnerk Schünemann. *Modell zur Bewertung der Herstellkosten von Lithiumionenbatteriezellen*. Sierke, 1st edition, 2015. ISBN 978-3-86844-704-0.

Naresh Susarla, Shabbir Ahmed, and Dennis W. Dees. Modeling and analysis of solvent removal during Li-ion battery electrode drying. *Journal of Power Sources*, 378(August 2017): 660–670, 2018. doi:10.1016/j.jpowsour.2018.01.007.

John C. Tannehill, Dale A. Anderson, and Richard H. Pletcher. *Computational Fluid Mechanics and Heat Transfer*. Taylor & Francis, Washington, 2nd edition, 1997. ISBN 0070503281.

Matthias Thomitzek, Felipe Cerdas, Sebastian Thiede, and Christoph Herrmann. Cradle-to-gate analysis of the embodied energy in lithium ion batteries. In *Procedia CIRP*, volume 80, pages 304–309, 2019. doi:10.1016/j.procir.2019.01.099.

Chris Yuan, Yelin Deng, Tonghui Li, and Fan Yang. Manufacturing energy analysis of lithium ion battery pack for electric vehicles. *CIRP Annals - Manufacturing Technology*, 66(1): 53–56, 2017. doi:10.1016/j.cirp.2017.04.109.

Complementing existing CHP plants with pyrolysis and gasification to produce liquid biofuels

Elena Tomás-Aparicio[1], Chaudhary Awais Salman[2], Erik Dahlquist[2], Selman Ahmeddin[2], Hussein Abdirizak Mohamed[2]

[1]Mälarenergi AB, Sweden elena.tomas-aparicio@malarenergi.se
[2]School of Business Society and Engineering, Division of Automation in Energy and Environmental Engineering, Mälardalen University, Sweden, {chaudhary.awais.salman, erik.dahlquist, san15006@student, hmd16001@student}@mdh.se

Abstract

In Northern Europe there are many CHP plants operating with biomass and waste as fuel. As more wind and solar power is introduced the operating hours of these plants is reduced and thereby also the capital burden is distributed on fewer annual hours. At the same time there is a strong request to replace fossil oil by renewable alternatives for many different purposes. Here biomass and waste are the major resources available to produce liquid or gaseous bio-based products at existing CHP plants. In this study we have simulated system solutions to identify energy and material balances as well as rough economic figures. The products assumed are primarily fuels like diesel, hydrogen and methane, but also other organic compounds can be considered. Today PREEM and St1 are planning large scale production of primarily bio-diesel, or HVO, where liquid products from both pulp and paper industry and CHP plants will be suitable feedstock. The study includes a comparison between hydrogen production in gasifiers to electrolysis, and even a combination of these as oxygen from the electrolyser can be used for the gasification, to avoid ballast of nitrogen in the product gas. The study aims to identify optimal solution under different conditions with respect to both electricity and raw material costs, as well as capital cost and operating hours.

Keywords: CHP, pyrolysis, gasification, HVO, liquid bio-fuel, hydrogen.

1 Introduction

Pyrolysis has been studied as a method to convert solid biomass into liquid bio-oil. A certain percentage of the biomass is converted to bio-oil and gaseous compounds, while a residue of mainly carbon is also produced. At Brista (Stockholm Exergi) an interest has been to produce bio-char primarily (Jonsson, 2016) while the bio-oil is the primary product in e.g. Joensuu pyrolyser plant (Joensuu, 2013). This is considered one of the first large scale pyrolisers for converting biomass to bio-oil. The production is around 50 000 ton per year. At Setra in Gävle, Sweden, saw dust will be converted to bio-oil as well, with a capacity of approximately 30 000 ton per

year (Setra, 2019). The goal is to have the plant up and running at the end of 2021. Here PREEM is one of the part owners. PREEM has a plan to produce 3 million m^3 bio-diesel and other bio-fuels per year by 2030 being this is a driving force for pyrolysis of biomass in Sweden. Already today PREEM is producing all their diesel with 30% HVO (hydrogenated vegetable oils) including one third tall oil and two thirds vegetable oil, mostly palm oil. At Chalmers University tests have been performed with pyrolysis (2 MW) in the G-valve of an 8 MW CFB (Circulating Fluidized Bed) boiler. Approximately 70 % of the biomass has been converted to gases, while the remaining solid fraction is passing down into the CFB bed, where the solids are combusted (Larsson et al, 2013). Most of the gases are condensed into liquids, which is an energy rich mixture of hydrocarbons. At Gobi gas a demonstration plant is operated to convert solid biomass into hydrocarbons using the FT (Fischer-Tropsch) process (Larsson et al, 2018). The product then can be used to replace fossil hydrocarbons. In Gussingen a steam-based gasification has been demonstrated in a CHP plant to produce a nitrogen free gas with high heating value (12 MJ/Nm^3). This was reported in Rauch et al (2004). An alternative route can also be to utilize black liquors from pulp and paper industry, and especially the tall oil. Tall oil has been refined at Sunpine in Piteå followed by hydrogenation and distilled at PREEM oil refinery in Gothenburg. Here the oil is reacted with hydrogen to get a product equal to the fossil oil used to produce diesel and kerosene. From the previous work it can be concluded that CHP plants can be used also for production of liquid bio-fuels. In this system study we are looking for the possibility to convert solid biomass to bio-oil by integration to existing CFB-boilers at Malarenergi AB's CHP, but principally any FB-boiler could be converted in a similar way.

2 System Study

The study investigates the integration of a pyrolysis reactor and a combination with a gasifier, alternative with an electrolyser to produce H$_2$, hydrogen, and also O$_2$, oxygen, which can be used as complement to air in the combustion and gasification in a CFB boiler with the

capacity of 150 - 180 MW$_{th}$ operating on biomass or organic waste. We then study complementing this with only a pyroliser or a combination with also a gasifier.

The sensitivity analysis includes the fluctuations in electricity price for purchased electric power, the price for biomass and waste as well as the annuity for the capital cost. Different operational modes are then evaluated assuming different demand of heat and electricity over a year in order to investigate reasonable operating hours for production of the liquid or gaseous fuel. Here also different products are assumed like pyrolysis liquid for HVO production (hydrogenated vegetable oil), methane and hydrogen. Hydrogen can be utilized either for refinement of the pyrolysis liquid directly or extracted and sold as a product for different applications like fuel to fuel cells or similar.

The system alternatives then become:

1. CFB with pyroliser (pyrolysis liquid as it is to a refinery).
2. CFB with gasifier (for hydrogen production primarily).
3. CFB with pyroliser combined with gasifier and gas upgrading with separation system.
4. CFB with pyroliser combined with electrolyser.
5. CFB with pyroliser combined with gasifier and gas upgrading with separation system plus electrolyser.

The system design with dimensioning of different equipment to balance the 180 MW thermal capacity of the CFB boiler is carried out. Thereafter a rough cost estimate is made for the investment cost and maintenance cost for the different solutions. The annual costs for each system design is estimated. Concerning dual bed gasifiers Lundberg et al (2018) have presented conversion measured in a pilot plant and discussed scale up of this.

The assumptions for some different possible scenarios are based on heat demand for previous years. Year 2016 when the temperature dropped to –20°C for several months, 2010 which was a harsh winter generally with much snow and 2019 when temperature has been quite high most of the time. On the other end we have summer 2018 when the temperature was 30°C and higher for two months. For "extreme periods" with very high or low temperature we look at hourly values, while normally monthly averages. From this we calculate energy demand with respect to heat, cooling and electricity and make a diagram showing the demand as a function of hours for each year.

We also look for the price of electricity, heat and cooling over the year for these years but also simulate other prices that might be expected in the future during different time periods. Here we look for how much of the total electricity production that comes from wind and solar power today and from this try to predict future

probable price. When there is a lot of wind power we can expect low electricity price and then electrolyzes may be the best alternative for hydrogen production, and summertime solar power will be in surplus

The calculations are plotted in a diagram where we can compare cost of product as a function of hours per year, fuel cost and electricity cost. The products assumed are H$_2$ and HVO.

3 Simulations

The model is primarily energy and mass balances for the different equipment complemented with chemical reactions for combustion, gasification and pyrolysis as well as electrolysis.

$$\frac{\partial m_i}{\partial t} = (\dot{m}_{in,i} - \dot{m}_{out,i}) \tag{1}$$

$$\frac{\partial m_i}{\partial t} = k\,(T) * m_{0,i} \tag{2}$$

$$\sum_i^n m_{i,in} x_{i,in} = \sum_i^k m_{i,out} x_{i,out} \tag{3}$$

$$\frac{\partial Q_i}{\partial t} = (\dot{m}_{in,i} - \dot{m}_{out,i}) * C_{p,i} * \Delta T + \Delta H * \frac{\partial m_i}{\partial t} \tag{4}$$

Here $\dot{m}_{in,i}$ is flow in (kg/s) of component i and $\dot{m}_{out,i}$ is flow out. k (T) is a reactivity constant for a given reaction, Q_i is the energy in kWh or kJ. $C_{p,i}$ is the heat capacity and ΔT is the temperature difference. ΔH is the heat released (or taken up) during the reaction.

Figure 1. System with CFB boiler, pyrolyser, gasifier and electrolyser.

The fluidization of sand was determined from the balance between gas flow and buoyancy forces versus gravity forces. For full combustion this forms only CO2

and H2O. For the gasification we get different composition depending on the relative oxidation (how much O2 is added in comparison to what is needed for 100% combustion), capacity (ton DS/m²h) and temperature. The gas composition for two different cases are given in Table 4 (Dahlquist et al, 2018). The energy balance in the CFB boiler is given in Table 1 and for the pyrolysis in the G-valve in Table 2. The boiler considered is recycled wood CFB boiler at Malarenergi with approximately 150 MW fuel feed. Part of this is used to heat sand which is passing over to the G-valve to heat fuel injected also there. This heat together with partial combustion is used as input to pyrolyse bio-mass and waste.

Table 1. Energy balance for the CFB boiler

Alternative	1
Fuel to CFB boiler	
kgTS/s	7.7
MW HHV (21MJ/kg)	161.7
MW LHV (19 MJ/kg)	146.3
Fuel to G-valve- pyrolyzer	
kgTS/s	7.3
MW HHV	152.5
MW LHV	138
Air boiler	
m3/s	27.3
kg/s (1 m3 = 1,293 kg)	35.3
MW to heat air to 850 oC	29.3
Air pyrolyzer	
m3/s	3
kg/s	3.9
MW to heat air to 450 oC	1.7
Sand to cyclone	
kg/s	54.6
temp boiler	850
temp after pyro	450
MW to pyrolyzer from sand	17.9

Table 2. Energy balance for the pyrolyser.

Pyrolyzer	
MW from combustion in pyrolyzer	
MW to heat fuel to 450 oC	2.6
Losses	
%	11.5
MW (incl drive pyrolyzis)	17.5
MW to drive pyrolyzis 1123 kJ/kg	8.2
Pyrolyzate liq	
%	64.6
MW	98.5
kg/s	4.7
kg/h	16884
Gas from pyrolyzis	
%	7.1
MW	10.8
Solid residue to CFB boiler from G-valve	
%	16.8
MW	25.6
Sum MW from pyrolyzer to boiler	36.4

3.1 Pyrolyser

Pyrolysis demands 8.2 MW while the heat from sand from the cyclone is 17.9 MW. Total losses are estimated to be 17.5 MW, assuming same proportion as in the small pilot plant. In reality losses can be reduced in a larger plant. Heat in gas and solids that are fed back to the CFB boiler are 36.4 MW solid organic residues and 10.8 MW gaseous compounds. The heating value in the pyrolysis liquid is 98.5 MW which correspond to 16.9 ton/h. Pyrolysis balances have been given in e.g. Neves et al (2015), where different conditions have been investigated. In our example we have used the balance presented by Atsonios et al. (2015) which is quite typical.

1 kg bio-oil may contain 46.7 mol C, 62 mol H_2 and 23.6 mol O. To remove 23.6 mol O we then need 23.6 mol H_2. 4.69 kg liquid/s means 23.6 mol O/kg*4.69 kg/s = 110.7 mol/s.

3.2 Gasifier

We have used experimental data from our pilot gasifier using wood pellets with 1.2 ton DS/m², h at 42 % relative oxidation (EOR) and steam addition. The gasification temperature was 670℃. The gas composition was 8.4% H_2, 10.9 % CO, 2.4% CH_4, 17.8 %CO_2, 14.9 % H_2O and 45.7% N_2. In reality the water content is quite uncertain as an unknown amount was condensing in the sampling system. We have made two assumptions. In the first case we just have used air as oxidation media and not treated the gas before the membrane separation. In the second case we have used oxygen instead of air and first condensed out water to 90% and then removed CO_2 to 90% using MEA. In the first case we can assume some H_2O and some CO_2 may be passed over with the H_2 through the membrane, while

in the second case the gas will be almost 100% H_2. In Table 3 we can see how the gas composition is changing over the different steps. In column 1 we have the raw gas after the gasifier. Then we have membrane filtrate where we have assumed 97% of the H_2 as clean gas in column 2 and residual gas going back to the CFB boiler in column 3. In column 4 we have the "raw gas" after condensing out water to 90% and removed CO_2 to 90% in a MEA scrubber or membrane filter with MEA at the back side. In column 5 we then have 100% H_2 and in column 6 the composition of the residual reject gas.

Table 3. The gas composition from raw gas to clean H_2 respectively residual gas for the two alternatives – gas separation in membrane filter directly respectively after separation of H_2O, CO_2 and not using air but oxygen from hydrolyzer.

| | Gas from gasifier | Membr separation direct 97% | | No N2 and remove H2O + CO2 to 90% | | | |
		Filtrate	Reject	Feed gas to membr	Filtrate	Reject	Raw gas without N
H2%	8.4	100	0.27	33.5	100	1.5	20.5
CO%	10.9		11.8	43.6		64.5	26.6
CH4%	2.4		2.6	9.6		14.3	5.9
CO2%	17.8		19.3	6.9		10.3	43.4
H2O%	14.9		16.3	6.4		9.4	3.7
N2%	45.6		49.7	0		0	0
	100		100	100		100	100

Calculations are now made for two different cases. In the first case it is actual gas composition from our pilot gasifier where we operated at 670 °C, 42 % relative oxidation and a capacity of 1.2 ton DS/m^2.h. The moisture content was 30% by addition of steam to wood pellets (Table 4 case 1). The second case is using a regression model from many gasification experiments in the pilot plant and assuming 35% relative oxidation, 2 ton DS/m^2.h, 800 °C, 30 % moist (Table 4 case 2).

Table 4. Gas composition and mol/kg DS for the different components for two different raw gas compositions.

| | Case 1 | | Case 2 | | |
| | Raw gas without N2 water condensed out | | Raw gas | Raw gas without N2 water condensed out | |
	%	mol/kg DS	%	%	mol/kg DS
H2%	20.5	17.7	20.2	40.7	41.6
CO%	26.6	14.9	12	24.2	18.2
CH4%	5.9	3.3	4.3	8.7	6.5
CO2%	43.3	24.4	11.8	23.8	17.8
H2O%	3.7	6.4	13.6	2.7	2.8
N2%	0	0	38.1	0	0
tot	100	66.7	100	100	86.9
H2+CO+CH4		39.2			72.8

The balances for the pyrolyser is taken from Atsonios et al (2015). We have noticed that for 4.67 kg pyrolizate/s we need to remove 110.7 mol O/s. For the first gas composition we produce 17.7 mol H_2/kgDS. (110.7 mol H_2/s)/ (17.7 mol H_2/kgDS) = 6.25 kg DS/s. If we can convert CO and CH4 to H2 as well we get 17.7 + 14.9 + 2*3, 3 = 39.2 mol H_2/kg DS. This would mean that 2.8 kg DS/s would be needed to the gasifier to cover the

reduction of Oxygen in the pyrolisate. For the second gas composition (case 2) we get 41.6 mol H_2/kg DS which gives a demand of 110.7/41.6 = 2.7 kg DS/s. If we convert also CO and CH4 to H2 through reaction with steam, we would get 72.8 mol H_2/kg DS and a demand of 110.7/72.8 = 1.52 kg DS/s. From this we can see that how the system is designed and how the gasifier is operated will have a significant effect on the capacity of the gasifier. With the second case the size would be 43 % compared to the first case for only H_2 utilized and 54 % if also CO and CH4 is converted to H_2.

If we have a HHV of the fuel at 21 MJ/kg we would need a gasifier with 1.52*21= 31.9 MW while in the case 6.25 kg DS/s*21 MJ/kg the capacity would be 131 MW. Cost estimates for the gasification and enrichment of H_2 has been presented in Naqvi et al (2017). We use the figure 70 M€ for a 180 MW gasifier plant and use the scaling factor 0.8, which gives 23.5 M€ for case 1 and 17.6 M€ for case 2. If we look at alternative 5 where we combine gasifier with electrolyser we get 13.5 respectively 10.1 M€ for the gasifiers but have to add the cost for electrolyser as well.

3.3 Electrolyser

Electrolysers use electricity to split water into H_2 and O_2. If we have a demand of 110.7 mol H_2/s it means that we also produce 55.4 mol O_2 (110.7 mol O). This can be used in the combustion, but especially in the pyrolyser and the gasifier if we combine with also this.

For the electrolyser we can assume the price to be some 500 €/kW at large scale up to the double for small scale electrolysers. Typically, one kg H_2 demand 58 kWhel for production, which means 0.017 kg H_2/h with one kW input electricity. For the demand 0.22 kg H_2/s then means a capacity of 0.221 kg H_2/s/(0.0000047 kg H_2/s)= 46 800 kW. For alternative 5 a 23 400 kW electrolyser is needed. We assume 0.12 as annuity (4%, 10y) and cost for electricity in the range 1-10 €cent/kWh$_{el}$. The prize of electricity will be dominating. When there is a surplus of electricity production we will be in the lower range of the span, while when there is a deficiency the prize will be in the upper range. This makes it more complicated to decide best technology. It would be good to have both gasifier and electrolyser and use the bio-gasifier when electricity prize is high, and electrolyser when it is low. If we assume a cost of 500 $/kW it means 446 €/kW in investment cost. This would give an investment cost of 20.9 M€ for the 46 800 kW unit and 10.5 M€ for the 23 400 kW unit.

4 Results and discussion

In Tables 5-7 we see the capacities for the different alternatives. We have assumed a biomass fuel price of 14 €/MWh, which is average today. For electricity we have assumed 10 €/MWh and 35 €/MWh, as we expect quite strong variation over the year, depending on the

balance between production and consumption. We also have looked at 5000 hours and 7000 hours of operations per year. All equipment is assumed having 20 years life time and 5% interest rate, giving annuity 0.081. Concerning the gasification and electrolyzes the figures are reasonable, while for the pyrolysis it is more of a "guestimate". The cost figures are taken from Naqvi et al (2017). There already are vessels for pyrolysis at the CFB boilers, the G-valves, but addition is needed with fuel feeder, gas outtake and gas condensation. As the reactor for H_2 with pyrolysis liquid is principally same for all alternatives except for alternative 1, it has not been included. We have assumed the value of the pyrolysis liquid to be 70% of the refined (reacted with H_2) for case 1. For case 2 we have just assumed same value for H_2 as for pyrolysis liquid. For case 2 and 3 we have added a cost for membrane separation for H_2 enrichment of 10 M€, and for case 5 half of this. These figures are taken from Naqvi et al (2017).

Table 5. 5000 hours per year, biofuel cost 14 €/MWh, electricity 10 €/MWh, annuity 0.081.

5000 h/year	Case				
Total cost incl capital/y	1	2	3	4	5
Capital cost(0.081) M€	3.2	2.7	6	4.9	5.5
Fuel 5000 h/y 14 €/MWh	13.9	12.3	19.4	17.2	12.3
Cost el 10€/MWh				2.3	1.2
Total cost per year	17.1	15	25.4	24.4	18.9
Income pyroliq	17.2	24.6	24.6	24.6	24.6
Gross benefit	0.12	9.7	´-0.7	0.2	5.7

Table 6. 7000 hours per year, biofuel cost 14 €/MWh, electricity 10 €/MWh, annuity 0.081.

7000 h/y el 10 €/MWh	Case				
Total cost incl capital/y	1	2	3	4	5
Capital cost(0.081) M€	3.2	2.7	6	4.9	5.5
Fuel 7000 h/y 14 €/MWh	19.4	19.4	19.4	19.4	17.2
Cost el 10€/MWh				3.3	1.6
Total cost per year	22.6	22.1	25.4	27.6	24.3
Income pyroliq	24.1	34.5	34.5	34.5	34.5
Gross benefit	1.5	12.4	9.1	6.9	10.2

Table 7. 7000 hours per year, biofuel cost 14 €/MWh, electricity 35 €/MWh, annuity 0.081

7000 h/y, el 35€/MWh	Case				
Total cost incl capital/y	1	2	3	4	5
Capital cost(0.081) M€	3.2	2.7	6	4.9	5.5
Fuel 7000 h/y 14 €/MWh	19.4	19.4	19.4	19.4	17.2
Cost el 35€/MWh				11.6	5.8
Total cost per year	22.6	22.1	25.4	35.9	28.4
Income pyroliq	24.1	34.5	34.5	34.5	34.5
Gross benefit	1.5	12.4	9.1	´-1.4	6.1

Table 8. 7000 hours per year, biofuel cost 14 €/MWh, electricity 100 €/MWh, annuity 0.081

7000 h/y, el 100 €/MWh	Case				
Total cost incl capital/y	1	2	3	4	5
Capital cost(0.081) M€	3.2	2.7	6	4.9	5.5
Fuel 7000 h/y 14 €/MWh	19.4	19.4	19.4	19.4	17.2
Cost el 35€/MWh				33	16.5
Total cost per year	22.6	22.1	25.4	57.3	39.1
Income pyroliq	24.1	34.5	34.5	34.5	34.5
Gross benefit	1.5	12.4	9.1	´-22.9	´- 4.7

From the results we can see that for 5000 hours operations the alternative with a smaller gasifier and a smaller electrolyser looks like a good alternative. The electrolyser is then used to produce both H_2 and O_2. The O_2 is used in the gasifier, and thereby giving a better gas from the gasifier, without N_2 from the air. For 7000 hours most cases are economic at low electricity price, but especially alternative 4 becomes very unprofitable with the higher electricity price. As we don´t have very accurate figures for the value of pyrolysis liquid of different qualities, and neither for large amounts of H_2, the results are more giving relative impact of different alternatives, but clearly show the importance of cost for both biomass and electricity. If we could use waste instead of biomass, we would have an income of 12 €/MWh instead of a cost for the better biomass of 14 €/MWh, but instead there would be an issue of what quality the pyrolysis liquid would have, if there is e.g. halogens from PVC in the plastic fraction.

From the analysis we can see that the span with respect to economy is high for the different alternatives. This is due to high difference in prize for different fuels both right now, but certainly also in the future. The capital cost also has gone down a lot last years for electrolysers and probably will proceed going down if the technology will be used much more frequently than in the past. We also can see that the sizing of especially the gasification plant will depend a lot on how much oxygen there will be in the liquid product after the pyrolysis. From literature we can see that the amount of oxygen varies between some 20 to 50 wt. %. If we can utilize both H_2 and CO for reduction of O in the liquid product it will also make a large difference compared to if only H_2 is used. Still, we have reliable figures on many process parts verified in both own experiments on gasification of different biomass and black liquors as well as reports from literature from experiment in pilot and demonstration plants, showing that each part of the processes are possible to get to work, although the complete systems have not been implemented yet.

5 Conclusions

The study shows that different system solutions can be the most economic depending on the conditions. If the price for pyrolysis liquid is good even without hydrogenating it, case 1 may be interesting. If there is a strong demand for H_2 as such, case 2 is feasible. If the value of the pyrolysis liquid is much higher after hydrogenation the addition of the gasifier make sense as in case 3. If the electricity price is low case 4 is good. If the electricity price is varying case 5 make sense, and also it is positive that the O_2 produced in the electrolyser can be utilized in the gasification. Thereby we get a N_2 free gas which is much easier to handle than gas with high amount of N_2. The cost figures are of varying quality and thus next step will be to do more detailed cost estimates and more detailed design. Here we will

utilize the simulation model we have developed to give the possibility to do more sensitivity analysis.

Acknowledgements

References

S. Ahmeddin and H. Mohamed. *Techno-economic evaluation for pyrolysis and gasification process integration with CHP plants to produce biofuels* (Master's thesis), 2020. http://urn.kb.se/resolve?urn=urn:nbn:se:mdh:diva-49014.

K. Atsonios, K. Panopoulos and E. Kakaras. Biomass fast pyrolysis energy balance of a 1kg/h test rig. *International Journal of Thermodynamics, 267,* 2015. doi: 10.5541/ijot.5000147483.

E. Dahlquist, M. Naqvi, E. Thorin, J. Yan, K. Kyprianidis and Philip Hartwell. Modeling of Wood Gasification in an Atmospheric CFB Plant. In *Proceedings of SIMS 2018, September 26 – 27, 2018, Oslo,* 872-877, 2018. doi:10.3384/ecp17142872.

Fortum Corporation. Bio-oil plant in Juensuu, 2013. https://www.fortum.com/media/2013/11/fortums-bio-oil-plant-commissioned-joensuu-first-its-kind-world.

E. Jonsson. *Slow Pyrolysis in Brista: An Evaluation of Heat and Biochar Production in Sweden* (Dissertation), 2016. http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-188601.

A. Larsson, M. Seemann, D.Neves and H. Thunman. Evaluation of Performance of Industrial-Scale Dual Fluidized Bed Gasifiers Using the Chalmers 2–4-MWth Gasifier. *Energy Fuels,* 27:6665–6680, 2013. doi: 10.1021/ef400981j.

A. Larsson, I. Gunnarsson and F. Tengberg. The GoBiGas Project Demonstration of the production of Biomethane from Biomass via Gasification. Göteborg Energi, 2018. 10.13140/RG.2.2.27352.55043.

L. Lundberg, D. Pallarès and H. Thunman. Upscaling Effects on Char Conversion in Dual Fluidized Bed Gasification. *Energy Fuels,* 32, 2018. doi: 10.1021/acs.energyfuels.8b00088.

M. Naqvi, E. Dahlquist and J. Yan. Complementing existing CHP plants using biomass for production of hydrogen and burning the residual gas in a CHP boiler. *Biofuels*, 8:1-9, 2016. doi:10.1080/17597269.2016.1153362.

D. Neves, H. Thunman, A. Matos, L. Tarelho and A. Gómez-Barea. Characterization and prediction of biomass pyrolysis products. *Fuel and Energy Abstracts*. 37:611-630, 2011. doi:10.1016/j.pecs.2011.01.001.

R. Rauch, H. Hofbauer, K. Bosch, I. Siefert, C. Aichernig, H. Tremmel, K. Voigtlaender, R. Koch and R. Lehner. Steam gasification of biomass at CHP plant in Güssing – status of the demonstration plant. In *Proceedings-2nd World Conference and Technology Exhibition on Biomass for Energy, Industry and Climate Protection*. 1687-1690. 2004.

Setra group. Setra och Preem först i Europa med förnybara drivmedel från sågspån. https://www.setragroup.com/sv/press/pressmeddelanden/2019/setra-och-preem-forst-i-europa-med-fornybara-drivmedel-fran-sagspan/. 2019.

Hybrid Model for Fast Solution of Thermal Synchronous Generator With Heat Exchanger

Khaled Aleikish, Madhusudhan Pandey, Thomas Øyvang, Bernt Lie

University of South-Eastern Norway, Porsgrunn, Norway, `Bernt.Lie@usn.no`

Abstract

Overheating of synchronous generators may lead to a shortened generator lifespan, thus strict constraints are imposed on their operation. A common constraint is to restrict the power factor of the generator to lie below, say, 0.86 overexcited. In some recent work, a dynamic thermal model of the generator with cooling heat exchanger has been developed; the idea is that this allows for better monitoring of generator temperature, while relaxing on the power factor constraint. The current model is only valid for an ideal case of constant heat capacity. In this work, the generator model is extended to allow for temperature dependence in heat capacity of water and air in the heat exchanger model. The consequence of this more realistic model, is that it is no longer possible to find an explicit, analytic solution of the heat exchanger model, and it is now necessary to instead solve numerically a two point boundary value problem for each time step in the differential equation solver.

It is shown that the effect of temperature dependence in the heat capacities has a noticeable effect on the solution of the model. The inclusion of on-line numeric solution of the heat exchanger model does, however increase the computation time of the thermal generator model by a factor of several thousand. Here, we instead study the possibility to solve the numeric heat exchanger model multiple times off-line, and then fit a regression model that gives a correction to the analytic solution. Both linear regression and nonlinear regression (neural network) is considered. Both types of regression models allow for a speed-up in the computation time of the thermal generator model of a factor of around 2000. In the computations, computer language Julia was used, with the DifferentialEquations and the Flux packages.

Keywords: linear regression, nonlinear regression, thermal model, machine learning, surrogate model, hybrid model.

1 Introduction

1.1 Background

Synchronous generators operated at high power factor face high currents, and potential overheating. To protect the generators from overheating, European hydropower generation limits the power factor to the range $[0.85, 0.95]$ overexcited, while in Norway the power factor should be less than or equal to 0.86 (Pandey, 2019). It is of interest to allow for more flexible constraints on the power factor to better handle sudden changes in consumption. To do this, it is necessary with good monitoring and control of generator temperatures; this possibility was discussed in (Øyvang, 2018).

1.2 Previous work

As a further study of the thermal generator model in (Øyvang, 2018), Lie[1] proposed a slightly more formal model, still assuming constant heat capacity in materials. This model was further studied in (Pandey, 2019); Pandey et al., 2019), where temperature dependences in heat capacities were introduced. Temperature dependence in heat capacities of air and water invalidates the analytic solution of the heat exchanger model, requiring the numeric solution of a static two-point boundary value (TPBV) problem for every time step in the time integration. The numeric solution of the TPBV problem is much slower than evaluating the analytic expressions (Lie[2]), leading to excessive computation time for model fitting, and on-line uses such as state estimation and advanced control. It is therefore of interest to study the development of simplified models/-surrogate models for the heat exchanger. A standard procedure would be to solve the TPBV heat exchanger model multiple times off-line, and then fit a regression model to the generated data. Data for surrogate models do not contain measurement noise *per se*, so an alternative to a regression model could be table look-up or interpolation. Still, regression models are often simpler, and there may be some "noise" due to inaccuracies in numeric computations. The problem of overfitting the data is relevant for surrogate models.

1.3 Overview of the paper

Instead of solving the nonlinear TPBV problem of temperature dependent heat capacities within the time integrator of the thermal synchronous generator model, data-driven surrogate models were developed for the nonlinear heat exchanger model as a correction to the analytic solution. The goal is that this hybrid heat exchanger model consisting in the analytic, ideal solution combined with the ex-

---

[1]Lie, B.: *Solution to Project* in course *FM1015 Modelling of Dynamic Systems* at University of South-Eastern Norway, 2018.

[2]Lie, B.: *Lecture Notes* in course *FM1015 Modelling of Dynamic Systems* at University of South-Eastern Norway, 2019.
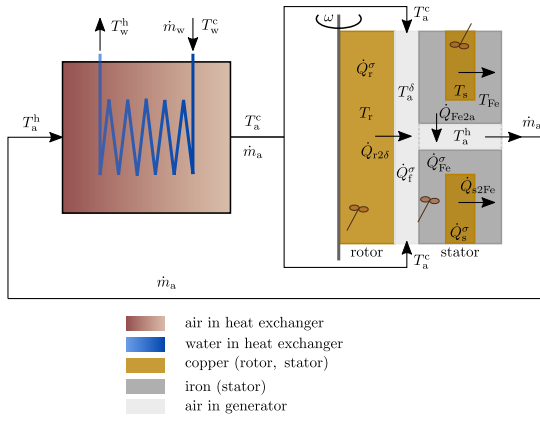
Figure 1. Thermal model of aircooled synchronous generator.

plicit surrogate correction, should give considerably faster solution time for the thermal generator model. Both linear regression and nonlinear regression (neural network) are considered, and validation is used to choose model order/avoid overfitting.

The paper is organized as follows. Section 2 describes the problem that arise from temperature dependence in the heat capacities and the impact on simulation time. Section 3 describes linear regression and the idea of validation, while Section 4 describes nonlinear regression. Results are discussed in Section 5, and some conclusions are given in Section 6.

2 Solution of the Counter-Current Heat Exchanger Model

A thermal model of a counter-current heat exchanger was developed in Lie[3].

The following explicit/analytic expressions were formed for the *effluent* temperatures of the tube side (T_e^t) and the shell side (T_e^s):

$$T_e^t = \frac{N_{St}^\triangle T_i^t \exp(-N_{St}^\triangle) + N_{St}^t T_i^s (1 - \exp(-N_{St}^\triangle))}{N_{St}^t - N_{St}^s \exp(-N_{St}^\triangle)} \quad (1)$$

$$T_e^s = \frac{N_{St}^\triangle T_i^s + N_{St}^s T_i^t (1 - \exp(-N_{St}^\triangle))}{N_{St}^t - N_{St}^s \exp(-N_{St}^\triangle)}. \quad (2)$$

Here, T_i^t and T_i^s are the *influent* temperatures of the heat exchanger. N_{St}^t and N_{St}^s are the *Stanton numbers* of the tube side and the shell side, respectively. Also, $N_{St}^\triangle$ is the difference in Stanton numbers, and is expressed as:

$$N_{St}^\triangle = N_{St}^t - N_{St}^s. \quad (3)$$

In addition, the temperature profile across the heat exchanger length can be obtained by substituting the explicit expression of T_e^s in the following equations:

$$T_t(x) = \frac{(N_{St}^t e^{(-N_{St}^\triangle)x} - N_{St}^s)T_i^t + (N_{St}^t - N_{St}^t e^{(-N_{St}^\triangle)x})T_e^s}{N_{St}^\triangle}$$

$$ \quad (4)$$

$$T_s(x) = \frac{(N_{St}^s e^{(-N_{St}^\triangle)x} - N_{St}^s)T_i^t + (N_{St}^t - N_{St}^s e^{(-N_{St}^\triangle)x})T_e^s}{N_{St}^\triangle}.$$

$$ \quad (5)$$

The Stanton numbers N_{St}^t and N_{St}^s for the tube side and the shell side, respectively, are given as

$$N_{St}^t = \frac{\mathscr{U}A_x}{\hat{c}_{p,t}\dot{m}_t} \quad (6)$$

$$N_{St}^s = \frac{\mathscr{U}A_x}{\hat{c}_{p,s}\dot{m}_s}, \quad (7)$$

where $\mathscr{U}A_x$ is the heat transfer coefficient, $\hat{c}_{p,t}$ and $\hat{c}_{p,s}$ are the specific heat capacities at constant pressure, and $\dot{m}_t$ and $\dot{m}_s$ are mass flow rates.

The previous analytic expressions are only valid for an ideal heat exchanger model with constant Stanton numbers. For a more realistic heat exchanger model, specifically, a model with temperature dependence in the heat capacities, a numeric solution is required. In this work, the two-point boundary value problem of the thermal model is solved numerically by utilizing the boundary value problem (BVP) solvers available in the DifferentialEquations package for Julia (Rackauckas and Nie, 2017).

In the thermal model of Lie[4], cold air is blown by a fan into the gap between the rotor and the stator, which cools the synchronous generator before returning to the shell side of the counter-current heat exchanger. There, the hot air is cooled by cold water passing in the tube side. In this work, temperature dependence is only considered in the heat capacities of air and water in the heat exchanger part of the model.

Temperature dependence in the heat capacities is often expressed as an empirical power series in T or as a polynomial in T (Murphy, 2020). In this work, polynomials were fitted to the experimental data in (Incropera et al., 2013) and compared with the empirical equations in (McBride et al., 2002). Figure 2 shows the comparison between the specific heat capacities of (Incropera et al., 2013) and the empirical equations of (McBride et al., 2002).

Furthermore, to study the impact of the temperature dependence of the specific heat capacities of air and water on the solution of the counter-current heat exchanger model, the following models are considered:

- Model 1: An ideal heat exchanger model, which is solved using the analytic expressions. The specific heat capacities of air and water are constant and equal 1.15 kJ/kg/K and 4.2 kJ/kg/K, respectively.

---

[3]Lie, B.: *Lecture Notes* in course *FM1015 Modelling of Dynamic Systems* at University of South-Eastern Norway, 2019.

[4]Lie, B.: *Solution to Project* in course *FM1015 Modelling of Dynamic Systems* at University of South-Eastern Norway, 2018.

Figure 2. A comparison between the specific heat capacities of (Incropera et al., 2013) and the empirical equations of (McBride et al., 2002).
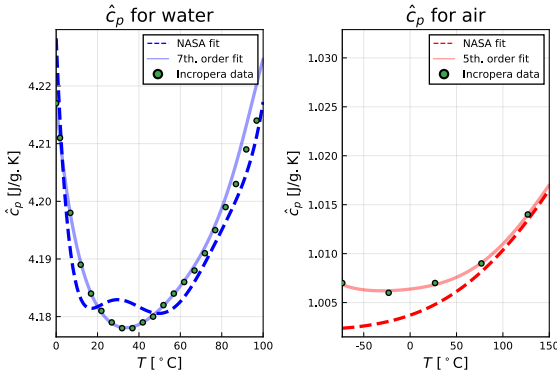
- Model 2: An ideal heat exchanger model, which is solved using the BVP solvers. The specific heat capacities are the same as in Model 1.

- Model 3: A non-ideal heat exchanger model with $\hat{c}_p(T)$, which is solved using the BVP solvers. The specific heat capacities are described by polynomials constructed from the experimental data in Incropera et al. (2013).

Figure 3 shows a comparison between the solution of Model 1 and the solution of Model 3, where the operating conditions are described in Table 1. Also, Table 2 shows the benchmark results (simulation time) for the three models.
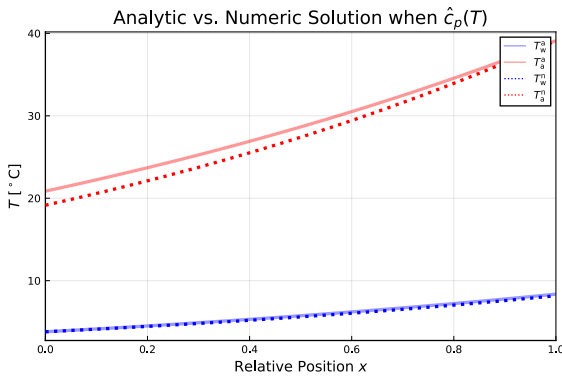
Figure 3. Analytic solution (solid lines) vs. numeric solution (dotted lines) when the specific heat capacities depend on temperature. For the ideal model, the specific heat capacities of air and water are constant, and equals 1.15 kJ/kg/K and 4.2 kJ/kg/K, respectively. For the numeric solution, the specific heat capacities are described by polynomials constructed from the experimental data in (Incropera et al., 2013).

Figure 4 shows a comparison between thermal generators model with heat exchanger Model 1 vs Model 3 at the operating conditions. Table 3 shows the benchmark results for the generator model when the three heat exchanger sub-models are used.

Table 1. Operating conditions of the heat exchanger model.

Symbol	Description	Value
	Inputs u	
T_w^c	Cold water temperature	3.8 °C
T_a^h	Hot air temperature	39.1 °C
$\dot{m}_w$	Water mass flow rate	53.9 kg/s
$\dot{m}_a$	Air mass flow rate	49.2 kg/s
	Parameters θ	
$\mathcal{U}A_x$	Heat transfer, air to water	44.46 kW/K

Table 2. Benchmark results for the heat exchanger model. Model 1 has median time $9\,\mu s$ and mean time $10.6\,\mu s$ on a given computer; times are scaled to unity at 1st row element.

Model	Median time	Mean time
Model 1	1	1
Model 2	443	451
Model 3	2 120	1 940

In this work, the following steps are taken to reduce the simulation time of Model 3 (the case of temperature dependence in the heat capacities):

- First, both Model 1 and Model 3 are solved for a variety of conditions (T_w^c, T_a^h, $\dot{m}_w$, and $\dot{m}_a$), and the results are stored in a data matrix.

- Next, a data-driven model is developed by linear/nonlinear regression. Here, the regression model is expressed as a correction term to the analytic expressions.

- Finally, the combination of the regression model and the analytic expressions forms explicit expressions that relate the influent temperature of water/air to the effluent temperature of water/air. Solving explicit expressions is much faster than solving a nonlinear two-point boundary value problem numerically.

3 Linear regression

As described in (Gujarati, 2019), given a data set $\{y_i, x_{1i}, x_{2i}, \ldots, x_{ki}\}_{i=1}^n$ of n observations, a generic linear regression model can be expressed as:

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + \varepsilon_i, \quad i = \{1, 2, 3, \ldots, n\};$$
(8)

here, y is called the response variable. $x_1, x_2, \ldots, x_k$ are called the regressor variables. $\beta_1, \beta_2, \ldots, \beta_k$ are the regression coefficients or regression parameters. ε is called the disturbance term. The subscript i refers to the observation in the data set. Moreover, equation 8 represents a system of equations that can be stacked together and written in matrix notation as:

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon,$$
(9)

where $\mathbf{y}$ and ε are $n \times 1$ column vectors, and β is a $k \times 1$ column vector. $\mathbf{X}$ is the design matrix of size $n \times k$. In this work, the possible regressors are:

Analytic vs Numeric Solution of Air-Cooled Synchronous Generator with Non-Ideal Heat Exchanger Model
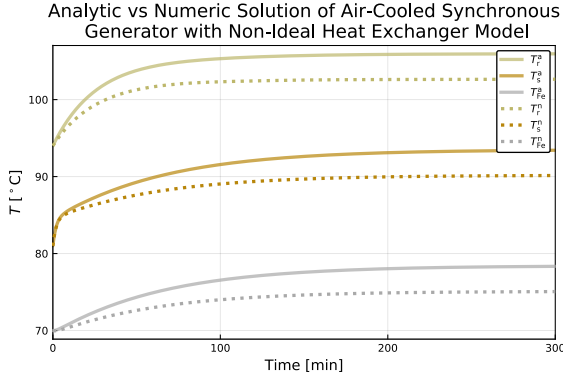
Figure 4. Analytic solution (solid lines) vs. numeric solution (dotted lines) of the thermal generator model when the specific heat capacities depend on temperature.

Table 3. Benchmark results for the thermal model of an air-cooled synchronous generator. Model 1 has median time 5.4 ms and mean time 6.6 ms on a given computer; times are scaled to unity at 1st row element.

Model	Median time	Mean time
Model 1	1	1
Model 2	451	363
Model 3	1 820	1 480

1. The hot water temperature of the analytic solution ($T_w^{h, A}$)

2. The cold air temperature of the analytic solution ($T_a^{c, A}$)

3. Water mass flow rate ($\dot{m}_w$)

4. Air mass flow rate ($\dot{m}_a$)

Also, the response variables are the effluent temperatures ($T_w^{h, N}$ and $T_a^{c, N}$) of the numeric solution when $\hat{c}_p(T)$. Since the number of response variables $m > 1$, then equation 9 is expressed as:

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon; \qquad (10)$$

here, $\mathbf{Y}$ and ε are $n \times m$ matrices, and β is a matrix of size $k \times m$. $\mathbf{X}$ is a design matrix of size $n \times k$. In this work, the columns of the design matrix are not the regressors themselves, but a polynomial of the regressors, which includes an intercept and cross-product terms. Using the method of ordinary least squares, an *estimate* of the regression coefficients can be obtained by:

$$\hat{\beta} = (\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T}\mathbf{y}. \qquad (11)$$

Then, the regression function is expressed as:

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\beta}; \qquad (12)$$

here, $\hat{\mathbf{y}}$ is the predicted value of $\mathbf{y}$ for a given design matrix. The regression function predicts the effluent temperatures of the non-ideal heat exchanger model given the effluent temperatures of the ideal model.

It is of interest to have a measure of *goodness-of-fit*, i.e., how well $\hat{y}$ predicts y. A simple predictor is the mean $\hat{y} = \bar{y}$, with uncertainty given by the sample variance

$$\sigma_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}; \qquad (13)$$

$\bar{y} = \sum_{i=1}^n y_i / n$. With predictor $\hat{y}_i = \sum_{j=1}^k \hat{\beta}_j x_{j,i}$, the variance is

$$\sigma_\varepsilon^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 / n \qquad (14)$$

with $\varepsilon_i = y - \hat{y}_i$. Two measures of model quality are the Root Mean Square Error (RMSE)[5]

$$\text{RMSE} = \sqrt{\sigma_\varepsilon^2} = \sigma_\varepsilon = \sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / n}, \qquad (15)$$

and the Coefficient of Determination R^2 (Gujarati, 2019)

$$R^2 = 1 - \sigma_\varepsilon^2 / \sigma_y^2. \qquad (16)$$

For a "perfect" model, $R^2 \to 1$, while for a "poor" model (the mean), $R^2 \to 0$.

The sample variance σ_y^2 is biased because Eq. 13 uses estimated mean $\bar{y}$; estimate σ_y^2 has one degree of freedom. The unbiased sample variance s_y^2 is

$$s_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n - 1} = \frac{n}{n - 1}\sigma_y^2. \qquad (17)$$

The prediction error variance uses k estimated parameters β_j, thus has k degrees of freedom, and the corrected estimate s_ε^2 is

$$s_\varepsilon^2 = \frac{n}{n - k}\sigma_\varepsilon^2. \qquad (18)$$

Alternatively to measures RMSE and R^2, the Standard Error of Estimate (SEE) (Smith, 2015)

$$\text{SEE} = \sqrt{s_\varepsilon^2} = s_\varepsilon = \sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n - k)}. \qquad (19)$$

or the adjusted coefficient of determination $\bar{R}^2$ defined as

$$\bar{R}^2 = 1 - s_\varepsilon^2 / s_y^2 = 1 - \frac{n - 1}{n - k}\left(1 - R^2\right) \qquad (20)$$

are used. RMSE and SEE has the advantage of having the same unit as y.

When solving Model 1 and Model 3, the following cases were considered:

- Case 1: Constant mass flow rates and variable influent temperatures. In this case, the regressors are $T_w^{h, A}$ and $T_a^{c, A}$. Also, the models are solved over the range of $(4 - 30\,°C)$ for T_w^c, and over the range of $(40 - 100\,°C)$ for T_a^h, while the mass flow rates are kept constant at the values of Table 1.

[5]https://en.wikipedia.org/w/index.php?title=Mean_squared_error&oldid=959959954

- Case 2: Variable influent temperatures and mass flow rates. In this case, the regressors are $\dot{m}_a$, $\dot{m}_w$, $T_w^{h,A}$, and $T_a^{c,A}$. Also, the models are solved over the range of $(4-40\,^\circ\text{C})$ for T_w^c, and over the range of $(15-130\,^\circ\text{C})$ for T_a^h, and over the range of $(15-130\,\text{kg/s})$ for the mass flow rates.

In Fig. 5 and Fig. 6, $\bar{R}^2$ and SEE were used to find the order of the polynomial of the design matrix that gives the best fit. Also, it can be observed in the figures that the goodness-of-fit deteriorates due to overfitting as the model order increases beyond a 5th order for case 1, and beyond a 6th order for case 2.

Moreover, Fig. 7 and Fig. 8 show a comparison between the regression models and the numerical solution of the nonlinear two-point boundary value problem for case 1. Also, in Fig. 7, the regression surface and the data points overlap for the 5th order model, which indicates a good fit. On the other hand, in Fig. 8, the regression surface of the 12th order model does not match the data for some parts, which indicates a bad fit.
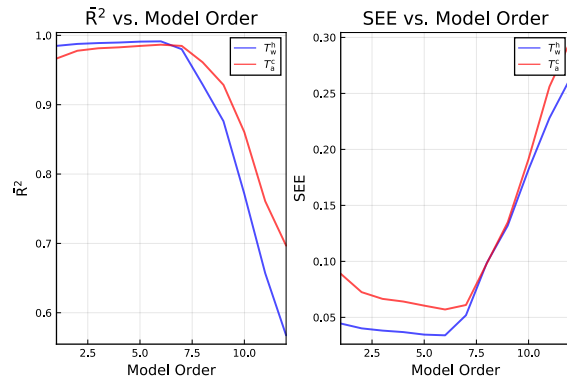
Similarly, Fig. 9 and Fig. 10 show a comparison between the regression models and the numerical solution of the nonlinear two-point boundary value problem for case 2. Also, in Fig. 9, the regression surface and the data points overlap for the 6th order model, which indicates a good fit. On the other hand, in Fig. 10, the regression surface of the 12th order model does not match the data for some parts, which indicates a bad fit.
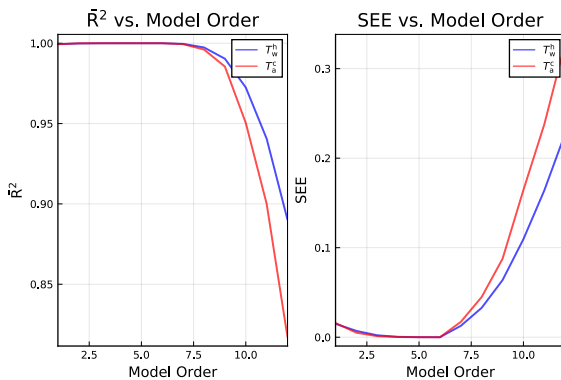

Figure 5. $\bar{R}^2$ and SEE of case 1. Left panel: The adjusted coefficient of determination vs. model order. Right panel: The standard error of estimate vs. model order.

However, the results so far are based on models that used the same data set for training and testing purposes, which is not a good indicator of the models' predictive ability. To evaluate a model's predictive performance and select the best fit model, validation is required.

In this work, the *Holdout* method was implemented to validate the models. The holdout method is one of the simplest cross-validation methods, where the data is split randomly into two data sets; a training set to estimate the model's parameters, and a test set to assess the performance of the model. The training data set is usually larger

Figure 6. $\bar{R}^2$ and SEE of case 2. Left panel: The adjusted coefficient of determination vs. model order. Right panel: The standard error of estimate vs. model order.
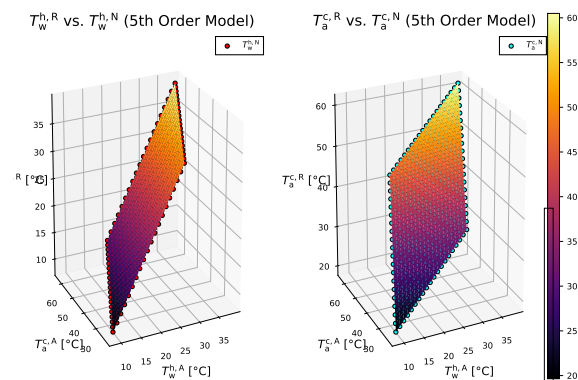
Figure 7. Case 1: A comparison between the regression model (the surfaces in the figure) and the numerical solution of the nonlinear two-point boundary value problem (the data points) for a 5th order regression model.

than the test set. Typically, the holdout method involves a single run. In this work, the results of multiple runs are averaged together to avoid misleading results. The validation results are presented in Table 4 for case 1, and in Table 5 for case 2.[6]

Finally, the best fit model of case 1 achieved an average RMSE of $7 \cdot 10^{-8}\,^\circ\text{C}$ and $2 \cdot 10^{-8}\,^\circ\text{C}$ for T_w^h and T_a^c, respectively. Also, the best fit model of case 2 achieved an average RMSE of $0.117\,^\circ\text{C}$ and $0.327\,^\circ\text{C}$ for T_w^h and T_a^c, respectively.

4 Nonlinear regression

Nonlinear regression of the counter-current heat exchanger model is implemented in Julia using the package Flux (Innes, 2018; Innes et al., 2018). Moreover, the nonlinear mapping between the analytic solution of the ideal counter-current heat exchanger model and the numeric solution of the non-ideal heat exchanger model (the case of temperature dependence in the specific heat capacities of air and water) is achieved using the logistic (also known

---

[6]Both SEE and $\bar{R}^2$ are rounded to the 5th digit after the decimal place.
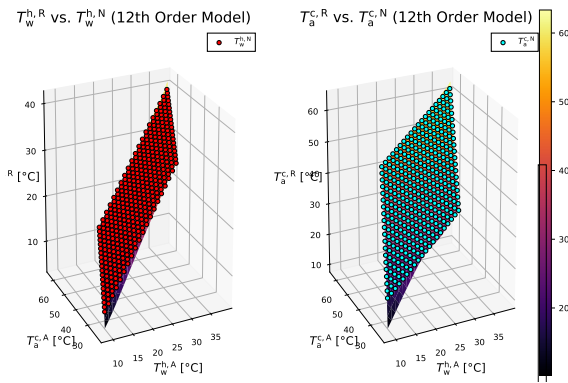
Figure 8. Case 1: A comparison between the regression model (the surfaces in the figure) and the numerical solution of the nonlinear two-point boundary value problem (the data points) for a 12th order regression model.
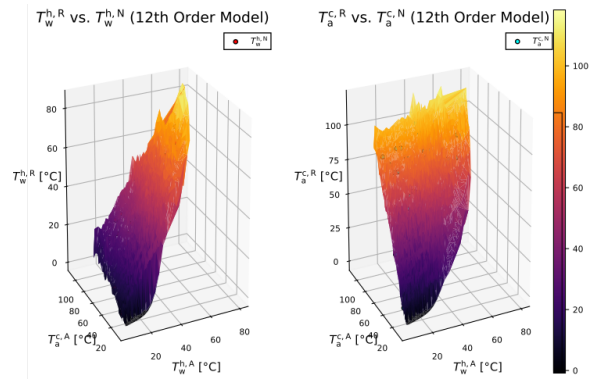
Figure 10. Case 2: A comparison between the regression model (the surfaces in the figure) and the numerical solution of the nonlinear two-point boundary value problem (the data points) for a 12th order regression model.
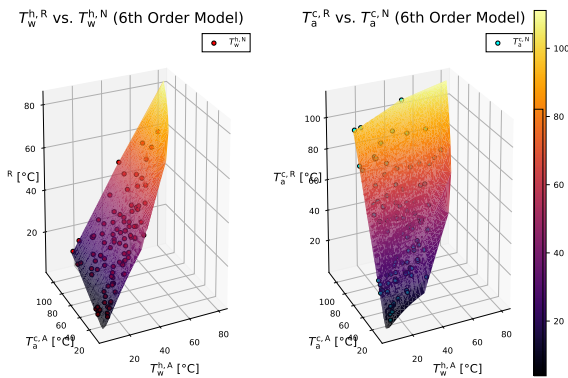
Figure 9. Case 2: A comparison between the regression model (the surfaces in the figure) and the numerical solution of the nonlinear two-point boundary value problem (the data points) for a 6th order regression model.

Table 4. Validation results of case 1.

Model order	Avg. SEE, °C		Avg. $\bar{R}^2$, −	
	T_w^h	T_a^c	T_w^h	T_a^c
1	0.01798	0.01905	0.99938	0.99944
2	0.00858	0.0061	0.99986	0.99994
3	0.00267	0.00149	0.99999	1.0
4	0.00056	0.00028	1.0	1.0
5	$7 \cdot 10^{-5}$	$4 \cdot 10^{-5}$	1	1
6	0.0001	0.00012	1	1
7	0.01823	0.02451	0.99947	0.99923
8	0.04843	0.06575	0.99642	0.99474
9	0.09783	0.13253	0.98625	0.97992
10	0.17664	0.27147	0.95805	0.91988
11	0.27841	0.4126	0.90428	0.832
12	0.41207	0.59617	0.81076	0.68251

as sigmoid) activation function (σ), which is introduced between two linear layers in the classical Feedforward Neural Network (FNN). Furthermore, the FNN is implemented using the description in (Lie, 2019) and Flux documentation. In this work, the nonlinear regression model is composed of two dense layers with the non-linearity (σ) between them as illustrated in Fig. 11.

In a similar manner to linear regression, the holdout method was implemented to select the dimension (out) in Fig. 11 that gives the best fit. The validation results are presented in Table 6 and Fig. 12, where RMSE was chosen as an indicator of goodness-of-fit. In Fig. 12, it is apparent that increasing the dimension (out) beyond 80 would result in a worse fit. Training of neural networks is sensitive to scaling of the data, and it is common practice to either *normalize* or *standardize* the data. Here, data X and Y have been normalized to $\tilde{X}$ and $\tilde{Y}$ in the range [0, 1], e.g.,

$$\tilde{X} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \qquad (21)$$

5 Results and Discussion

In linear/nonlinear regression, the counter-current heat exchanger model was solved for a variety of conditions (T_a^c, T_a^h, $\dot{m}_w$, $\dot{m}_a$, and thermal dependencies) to generate a data matrix. Then, validation was used to select the order of the polynomial and the dimension (out) that gives the best fit. Both case 2 of linear regression and nonlinear regression obtained a similar RMSE. Case 1 of linear regression achieved lower SEE and RMSE compared to case 2. However, case 1 is only valid for a constant mass flow rate, whereas case 2 is valid over a wide range of mass flow rates. This highlights the importance of selecting the appropriate ranges of T_w^c, T_a^h, $\dot{m}_w$ and $\dot{m}_a$ when solving the model. Selecting the appropriate ranges will result in a lower model order and a better fit.

In this work, the polynomial of the design matrix was generated using a user made function, which made the execution speed of the hybrid solution with linear regression relatively slower than the execution speed of the hybrid solution with nonlinear regression. Table 7 compares the execution speed of both solutions. To present a fair comparison between the linear and the nonlinear models, a 2nd

Table 5. Validation results of case 2.

Model	Avg. SEE, °C		Avg. $\bar{R}^2$, −	
order	T_w^h	T_a^c	T_w^h	T_a^c
1	0.0529	0.1059	0.9848	0.9665
2	0.0479	0.0864	0.9876	0.9777
3	0.0456	0.0795	0.9888	0.9812
4	0.0443	0.0771	0.9895	0.9825
5	0.0418	0.0731	0.9908	0.9844
6	0.0415	0.0695	0.991	0.9862
7	0.0686	0.0783	0.976	0.9828
8	0.1246	0.1246	0.9241	0.9582
9	0.1731	0.1774	0.8606	0.9192
10	0.2467	0.2587	0.7357	0.8396
11	0.3229	0.3556	0.5897	0.7253
12	0.4014	0.4455	0.447	0.6229
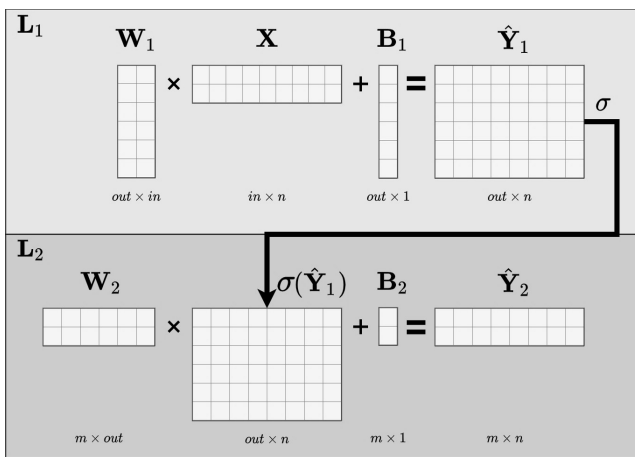
Figure 11. Feedforward neural network with two dense layers (L1 and L2) and an activation function (σ) in between them.

order linear model was hard-coded in the design matrix, which also provides a good enough prediction accuracy.

The simulation time of the final hybrid models, the combination of the correction terms of linear/nonlinear regression with the analytic expressions of equations 1 and 2, is presented in Table 8 for the heat exchanger model, and in Table 9 for the thermal model of an air-cooled synchronous generator. The hybrid solutions of the heat exchanger sub-model achieved similar execution speeds and are much faster than the numeric solution of the nonlinear two-point boundary value problem. Similarly, the simulation time of the thermal model of an air-cooled synchronous generator was significantly reduced by using the hybrid models.

6 Conclusions

In this paper, the thermal model of the counter-current heat exchanger that was developed in Lie[7], is extended with the case of temperature dependence in the specific heat ca-

---

[7]Lie, B.: *Lecture Notes* in course *FM1015 Modelling of Dynamic Systems* at University of South-Eastern Norway, 2019.
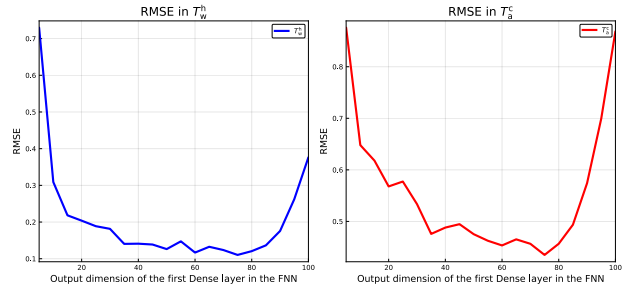
Figure 12. Validation results of nonlinear regression of the counter-current heat exchanger model. The Y-axis is the average RMSE for three models, each trained on a random sample for 10000 epoch. The best fit model achieved an average RMSE of $0.110\,°C$ and $0.435\,°C$ for T_w^h and T_a^c, respectively.

Table 6. Validation results of nonlinear regression.

Model	Avg. RMSE, °C	
order	T_w^h	T_a^c
5	0.7297	0.8757
10	0.309	0.648
15	0.2183	0.6179
20	0.2038	0.5679
25	0.1887	0.5774
30	0.1815	0.5336
35	0.1405	0.4758
40	0.141	0.4881
45	0.1387	0.4947
50	0.1263	0.4752
55	0.1474	0.4625
60	0.1168	0.4534
65	0.1326	0.4652
70	0.1237	0.4566
75	0.1104	0.435
80	0.1207	0.4567
85	0.1365	0.4936
90	0.1756	0.5742
95	0.2621	0.6995
100	0.3765	0.8686

pacities of air and water. The benchmark results showed a very long simulation time when solving the nonlinear boundary value problem numerically. To speed up the solution time, explicit data-driven models were developed using linear and nonlinear regression. Validation was used to select the order of the polynomial of the design matrix and the dimension of the layers in the FNN. However, the order of the polynomial and the dimension of the layers were not the only factors that impacts the accuracy of the models. The accuracy of the models was also affected by the generated data matrix on which the regression models were fitted. The generated data matrix should only contain informative data that relates to the model's objective. Moreover, the regression models were expressed as a correction term to the explicit/analytic ideal heat exchanger model. The hybrid heat exchanger model achieved a good enough accuracy and faster execution speed compared to

Table 7. The execution speed of the hybrid solution: linear regression vs nonlinear regression. Here, the polynomial of the design matrix was generated using a user made function, which made this comparison not fair. Linear-hybrid model has mean time $36\,\mu$s on a given computer; times are scaled to unity at 1st row element.

Model	Mean time
Hybrid solution (Linear Reg.)	1
Hybrid solution (Nonlinear Reg.)	0.042

Table 8. Simulation time for the heat exchanger model. Linear-hybrid model has mean time $1.5\,\mu$s on a given computer; times are scaled to unity for 1st row element.

Model	Mean time
Hybrid solution (Linear Reg.)	1
Hybrid solution (Nonlinear Reg.)	1.002
Numeric solution	14070

the numeric solution.

It should be observed that the hybrid heat exchanger model introduced here is based upon the assumption that there are no unknown model parameters in the heat exchanger. In other words: if we want to include the hybrid heat exchanger model in a model fitting procedure for the thermal generator model, we can not change the heat exchanger parameters, e.g., $\mathscr{U}A_{\text{x}}$ in Table 1 — that would require a re-training of the correction term.

Future work will include a study on how the correction term to the analytic heat exchanger can be re-formulated in order to allow for fitting model parameters such as $\mathscr{U}A_{\text{x}}$ without having to re-train the correction term. As part of this, the value of $\mathscr{U}_{\text{x}}$ will vary with flow rates. Next, the work reported in (Pandey et al., 2019) will be updated with this new thermal synchronous generator model with the more accurate heat exchanger model.

References

Damodar N. Gujarati. *Linear Regression: A Mathematical Introduction*. SAGE Publications, Inc, 2019. ISBN 978-1-5443-3657-2. doi:10.4135/9781071802571.

Frank P. Incropera, David P. Dewitt, Theodore L. Bergman, and Adrienne S. Lavine. *Principles of Heat and Mass Transfer*. International Student Version. John Wiley & Sons, Inc., Singapore, 7 edition, 2013. ISBN 968-0-470-64615-1.

Michael Innes, Elliot Saba, Keno Fischer, Dhairya Gandhi, Marco Concetto Rudilosso, Neethu Mariya Joy, Tejan Karmali, Avik Pal, and Viral Shah. Fashionable modelling with flux. *CoRR*, abs/1811.01457, 2018. doi:1811.01457.

Mike Innes. Flux: Elegant Machine Learning with Julia. *Journal of Open Source Software*, 2018. doi:10.21105/joss.00602.

Bernt Lie. Surrogate and hybrid models for control. In *Proceedings of the 60th Conference on Simulation and Modelling*, volume 170 of *Linköping Electronic Conference Proceedings*, pages 1–8, University of Västerås, Västerås, Sweden,

Table 9. Simulation time for the thermal model of an air-cooled synchronous generator. Linear-hybrid model has mean time $2.5\,$ms on a given computer; times are scaled to unity for 1st row element.

Model	Mean time
Hybrid solution (Linear Reg.)	1
Hybrid solution (Nonlinear Reg.)	1.1
Numeric solution	3920

August 2019. SIMS, Linköping University Electronic Press. doi:https://doi.org/10.3384/ecp201701.

Bonnie J McBride, Michael J Zehe, and Sanford Gordon. NASA Glenn Coefficients for Calculating Thermodynamic Properties of Individual Species. Technical Report NASA/TP–2002-21155, NASA, NASA Center for Aerospace Information 7121 Standard Drive Hanover, MD 21076, 2002. URL `http://gltrs.grc.nasa.gov/GLTRS`.

Kathleen E. Murphy. *Thermodynamics Problem Solving in Physical Chemistry: Study Guide and Map*. CRC Press, 2020. ISBN 978-1-00-003028-0.

Thomas Øyvang. *Enhanced power capability of generator unites for increased operational security*. PhD thesis, University of South-Eastern Norway, Faculty of Technology, Natural Sciences and Maritime Sciences University of South-Eastern Norway N-2018 Porsgrunn Norway, December 2018. ISBN: 978-82-7206-503-3 (print) ISBN: 978-82-7206-504-0 (online).

Madhusudhan Pandey. Model fitting and state estimation for thermal model of synchronous generator. mathesis, University of South-Eastern Norway, Porsgrunn, Norway, 2019.

Madhusudhan Pandey, Thomas Øyvang, and Bernt Lie. State estimation of a thermal model of air-cooled synchronous generator. In *Proceedings of the 60th Conference on Simulation and Modelling*, volume 170 of *Linköping Electronic Conference Proceedings*, pages 190–197, University of Västerås, Västerås, Sweden, August 2019. SIMS, Linköping University Electronic Press. doi:https://doi.org/10.3384/ecp20170190.

Christopher Rackauckas and Qing Nie. DifferentialEquations.jl — A Performant and Feature-Rich Ecosystem for Solving Differential Equations in Julia. *Journal of Open Research Software*, 5(15), 2017. doi:10.5334/jors.151.

Gary Smith. *Essential Statistics, Regression, and Econometrics*, chapter 10 - Multiple Regression, pages 301–337. Academic Press, second edition, 2015. doi:https://doi.org/10.1016/B978-0-12-803459-0.00010-8.

Recurrent Neural Network Based Simulation of a Single Shaft Gas Turbine

Hamid Asgari[1]   Emmanuel Ory[1]   Jari Lappalainen[2]

[1]VTT Technical Research Centre of Finland Ltd., Espoo, Finland,
{hamid.asgari,emmanuel.ory}@vtt.fi
[2]Semantum Oy, Finland,
jari.lappalainen@semantum.fi

Abstract

In this study, a model of a single shaft *gas turbine (GT)* is developed by using *artificial intelligence (AI)*. A *recurrent neural network (RNN)* is employed to train the datasets of the *GT* variables in *Python* programming environment by using *Pyrenn* Toolbox. The resulting model is validated against the *Test* datasets. Thirteen significant variables of the gas turbine are considered for the modelling process. The results show that the *RNN* model developed in this study is capable of performance prediction of the system with a high reliability and accuracy. This methodology provides a simple and effective approach in dynamic simulation of gas turbines, especially when real datasets are only available over a limited operational range and using simulated datasets for modelling and simulation purposes is unavoidable.

Keywords: gas turbine, modelling, simulation, artificial intelligence, recurrent neural network, black-box model

1 Introduction

Todays, *Artificial intelligence (AI)* plays a key role in the industrial world and has a meaningful and influential presence in many aspects of daily life. *AI* is an area of computer science that deals with performing human-like tasks. As one of the most popular applications of *AI*, one can refer to *Machine learning (ML)*. *ML* is a research area that deals with data analysis, and a variety of approaches for computers to learn from datasets when no specific algorithm exits. *ML* is used in many industrial systems such as gas turbines for pattern recognition, system identification, trouble shooting, sensor validation, condition monitoring, modelling, simulation, and performance prediction (Liu et al., 2020; Cisotto and Herzallah, 2018; Kiakojoori and Khorasani, 2016).

One of the main tools used in machine learning is *artificial neural network (ANN)*. *ANN*, as a subset of artificial intelligence, has been widely used over the past decades. The main idea for creating an *ANN* was to mimic the human brain as a simple model that was capable of solving complex problems in different scientific fields. As a brain-inspired data-driven model, *ANN* has shown to be a very good alternative to white-box approach for modelling and simulation of industrial systems and processes. It has been very useful for the researchers whose activities cover a wide range of system identification and simulation techniques. A recurrent neural network is a subset of *ANN* that is used to setup dynamic models of industrial systems.

Gas turbines are broadly employed, such as airplane engines. They are also widely used for power generation, especially in the areas that are far away from cities; where access to the national electrical grid is limited. Oil & gas fields, offshore platforms, ships, and marine infrastructures are examples of such places. The significant role of *GT*s in industry has strongly encouraged and motivated researchers to investigate new methodologies for modelling these complex systems. The desired models should be able to capture the system dynamics as accurately as possible. So far, many different experimental and analytical methodologies have been explored to reach this goal. The research in this area is included in both stationary (mainly located in power plants) and aero gas turbine engines (Asgari et al., 2012, 2013a).

ML-based techniques like *ANN* have shown the capability to predict dynamic behavior of *GT*s without having access to information about the system physics. Different *ANN-based* methodologies have already been investigated and developed in order to disclose complex nonlinear behavior of aero gas turbines (Agrawal and Yunis, 1982; Chiras et al., 2001, 2002; Ruano et al., 2003; Torella et al., 2003; Sarkar et al., 2012, 2013; Salehi and Montazeri, 2018; Ibrahem et al. 2019). These efforts have covered a variety of approaches such as *MLP (multi-layer perceptron)*, *NARMAX (nonlinear auto-regressive moving average with exogenous inputs)*, *NARX (nonlinear autoregressive exogenous model)*, *RBF (radial basis function)*, *BPNN (back propagation neural networks)*, and *BSNN (B-spline neural networks)*.

Besides, many studies have focused on *ANN-based* modelling of *industrial power plant gas turbines (IPGT)*. The results of these studies demonstrated the capability of *ANN* in capturing *IPGT* dynamics

(Lazzaretto and Toffolo, 2001; Kim et al., 2001, 2002; Ogaji et al., 2002; Basso et al., 2004; Jurado, 2005; Simani and Patton, 2008; Fast et al., 2008, 2009; Yoru et al., 2009; Tavakoli et al., 2009; Fast and Palmé, 2010; Palmé et al., 2011; Bartolini et al., 2011; Asgari et al., 2013b, 2014, 2015, 2016). *ANN* has also been used for control-oriented modelling of *GT*s (Asgari et al., 2017). *ANN* may be used for fault identification and warning generation with high reliability. Arriagada et al. (2003), Elashmawi et al. (2017, 2018), and Rahmoune et al. (2017) used novel *ANN-based* models for monitoring and fault detection of industrial gas turbines.

From the literature survey, many efforts have been made in the area of black-box modelling of gas turbines, with their own advantages and limitations. However, because of complex nonlinear dynamics of *GT*s, and their different applications, sizes and capacities, study in this area is still a challenging issue. Further research still need to be carried out to develop the *GT* models, and to optimize design and performance of these engines. The work reported in this publication can help to disclose details of the problems that cause unpredictable shutdowns, over-heating, and over-speed of *GT*s on industrial sites.

In this study, a black-box model of a single shaft gas turbine is modelled and simulated by using a recurrent neural network. The *RNN* is employed to train datasets of the gas turbine in *Python* programming environment. First, the performance of a typical gas turbine is briefly explained in *Section 2*. Then, in *Section 3*, a short description about recurrent neural networks is presented. *Section 4* describes modelling procedure including data acquisition, data processing, building *RNN* structure, code generation, and training process. The results are presented in *Section 5*. Finally, the output of the research is briefly discussed in *Section 6*.

2 Gas Turbines

A gas turbine is classified as an internal combustion engine that converts chemical energy to mechanical energy. It is widely used for industrial applications. *GT* may be connected to a generator, pump or compressor as the main driver to shape a turbo-generator, a turbo-pump, or a turbo-compressor respectively. Figure 1 illustrates the main components of a single shaft open-cycle constant-pressure gas turbine engine (Encyclopedia Britannica, Inc., 1999). Both the compressor and the turbine are installed on the central shaft and rotate together.

Operation of an industrial gas turbine is shown in Figure 2. According to this figure, the compressor receives air at point 1 and delivers hot compressed air to the combustion chamber (combustor) at point 2. The mixture of fuel and air is ignited inside the combustion chamber to form the hot gases at point 3. These hot gases pass through the turbine and rotate it. The output power

of the turbine provides the required energy for driving the compressor part, and the *GT* mechanical output.

Operation of a gas turbine is based on Brayton cycle. A temperature-entropy (T-S) framework-based of a standard Brayton cycle is illustrated in Figure 3 (Arabi et al., 2019). According to this figure, 1-2-3-4-1 shows the ideal cycle, while 1-2´-3´-4´-1 indicates the real cycle. In the real cycle, processes in both the compressor (1-2´) and the turbine (3´-4´) are non-isentropic and irreversible. In the ideal cycle, these processes are assumed isentropic. Processes 2-3 and 4-1 may be considered isobar, if pressure losses in combustor and air filters are neglected (Walsh and Fletcher, 1998).
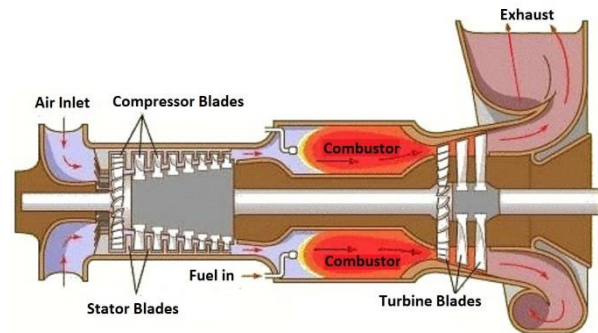

Figure 1. A typical single shaft gas turbine engine (Encyclopedia Britannica, Inc., 1999).
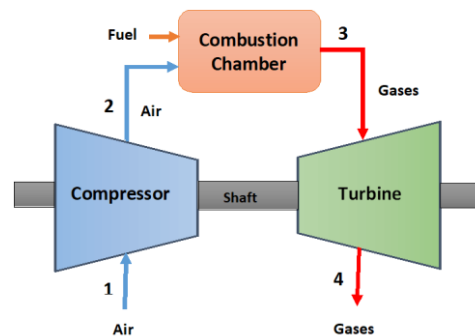

Figure 2. A schematic of a typical single shaft gas turbine.
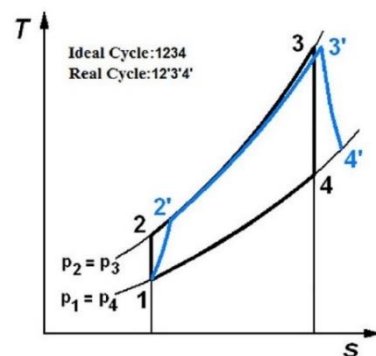
.

Figure 3. Ideal and real Brayton cycles in temperature-entropy diagram (Arabi et al., 2019).

3 Recurrent Neural Networks

Simulation of all industrial systems can be carried out by using white-box or black-box models. When knowledge and information about the physics involved in a system are quite enough and dynamic model equations are available, white-box models can be employed, and different simulation software such as *MATLAB* and *APROS* may be used for modelling of the system dynamics. *MATLAB* is a multi-functional programming platform designed and developed by *MathWorks* for numerical computing in engineering and scientific applications. *APROS* is also a multi-functional software for modelling and dynamic simulation of industrial processes, such as power plants and other energy systems (*Fortum* and *VTT*, 2020). Dynamic equations usually need to be simplified (e.g. linearized) because of their nonlinear and coupled natures. However, the simplification process may negatively affect the accuracy of the resulting model.

When the physics of a system is unknown or access to the relevant knowledge, dynamic equations and efficient solution techniques is limited, black-box modelling approach can be a good choice, and sometimes unavoidable. Black-box models can disclose the relationships among the system variables by employing input and output datasets. These datasets can be simulated or measured (experimental) data. Simulated data may be used when operational data are not available.

ANN is a class of black-box models that can be used for simulation of gas turbines and other industrial equipment (Tiumentsev and Egorchev, 2019). The structure of an *ANN* consists of input, hidden and output layers. Each *ANN* may have more than one hidden layer. The layers may have different number of artificial neurons. These neurons (units) are internally connected by transfer functions, which can be linear or nonlinear. Artificial neural networks are capable of learning the relationship among inputs and outputs of a system through an iterative training process. Each input into the neuron is associated with its own adjustable number, which is called weight. Weights are determined during the training process. The complexity of a system dynamics determines the number of hidden layers and their associated neurons. Figure 4 illustrates a typical *ANN* including three inputs, two outputs and one hidden layer with four neurons.

Among different *ANN* algorithms for static and dynamic modellings, *RNN* can be employed for the modelling of dynamic industrial systems. In a recurrent neural network, each layer has a recurrent connection. This enables *RNN* to propagate data forward and backward, from later processing stages to earlier ones, allowing the network to have an infinite dynamic response to the input data. As a universal approximator, *RNN* has shown excellent dynamic ability to deal with various input and output types for modelling and simulation of industrial systems.
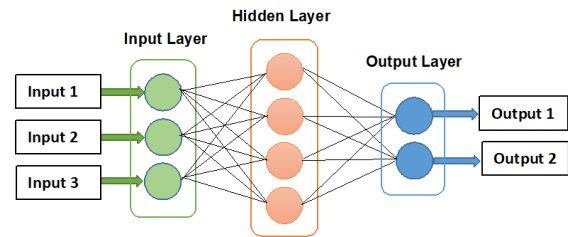
Figure 4. The structure of a typical artificial neural network with input, hidden and output layers.

4 Modelling Procedure

To approach an *RNN* model with a high accuracy for the gas turbine of this study, a variety of structures was considered. These structures were set up based on the data type, training algorithms, types of activation functions, number of hidden layers, number of neurons, and values of the weights and biases. The goal was to attain a structure with the high capability of accurate prediction of the *GT* dynamic behavior. Using the most effective *GT* variables as inputs and outputs is vital for building a reliable model. Data availability, system knowledge, and modelling objectives are fundamental factors that should be considered for selection of the *RNN* inputs and outputs.

4.1 Data Acquisition

The first step in *RNN* modelling is to obtain enough reliable datasets. A dynamic model of a low-power gas turbine, simulated in *Simulink/MATLAB* environment was employed for data generation (Asgari et al. 2013b). The relevant white-box model was already developed and verified against experimental datasets (Ailer et al., 2002). Totally, 3000 datasets were generated for 13 different *GT* variables. In this study, the purpose of a dataset is a vector with 13 single values, corresponding to 13 different *GT* variables. Therefore, 3000 datasets cover 3000 vectors like that.

The obtained datasets were categorized as inputs and outputs of the gas turbine engine according to Table 1 and Table 2 respectively. As it can be seen from the tables, fuel rate, ambient temperature, ambient pressure, and load were determined as four main gas turbine inputs, while temperature and pressure at different operational points (corresponding to the numbers in Figures 2 and 3), compressor pressure ratio, rotational speed, and gas turbine efficiency were considered as *GT* outputs. To be able to evaluate the model generalization, and to avoid over-fitting during the training process, the datasets were alternatively divided into the *Train* and *Test* groups. Half of the datasets (1500 out of 3000) were employed for training the *RNN*, and the remaining datasets were used to test the resulting model.

Table 1. Gas Turbine Input Parameters.

Variable	Symbol	Unit	Operational Range
Fuel rate	M_f	kg/s	[0.00367; 0.027]
Ambient temperature	T_{01}	K	[243.15; 308.15]
Ambient pressure	P_{01}	kPa	[60; 110]
Load	M_{Load}	N.m	[0; 363]

Table 2. Gas Turbine Output Parameters.

Variable	Symbol	Unit
Temperature at point 2	T_{02}	K
Temperature at point 3	T_{03}	K
Temperature at point 4	T_{04}	K
Pressure at point 2	P_{02}	kPa
Pressure at point 3	P_{03}	kPa
Pressure at point 4	P_{04}	kPa
Compressor pressure ratio	CPR	-
Rotational speed (number of revolutions)	N	1/s
Gas turbine efficiency	E_{GT}	-

4.2 RNN Structure

Since the gas turbine is a dynamic system, the aim of this research is to create a recurrent neural network, in which the output parameters of the current time-step depends on the output parameters in the previous one. In this study, an *RNN* model was designed with four inputs, one hidden layer, and nine outputs. The network was named *RNN 4-H-9* according to its structure (see Figure 5).
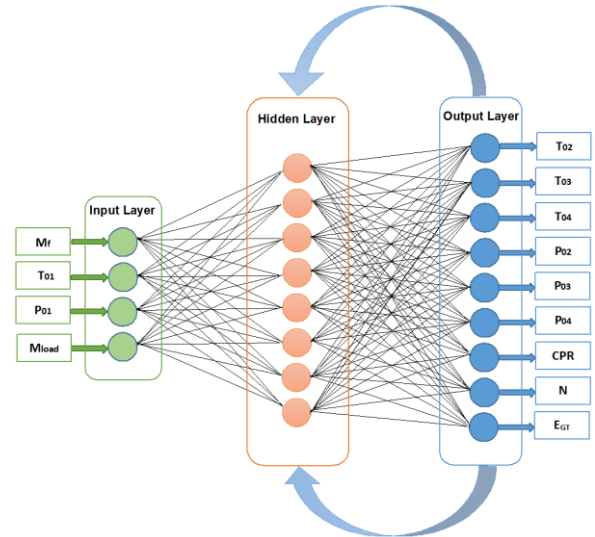
Figure 5. The structure of the recurrent neural network *RNN 4-H-9*.

4.3 Code Generation & Training Process

To achieve an *RNN* model with good generalization characteristic, details of the network structure should be determined as accurately as possible. For this purpose, a comprehensive computer code was developed in *Python* programming environment. *Python* is a high level, interpreted, and object-oriented programming language, firstly created and used by Guido van Rossum in 1991. The *Pyrenn* Toolbox, integrated in *Python*, was used for training and testing *RNN 4-H-9*. *Pyrenn* allows the creation of a wide range of *RNN* configurations. It employs *LM (Levenberg–Marquardt) algorithm* for training the network. *LM* is a second-order Quasi-Newton optimization algorithm, which is much faster than first-order methods such as gradient descent (Suzuki, 2011). The *RNN 4-H-9* model was trained by following a flexible programming code and assigning a combination of different values for the number of neurons in the hidden layer, the maximum number of iterations, the number of recurrent connections, and delay in the recurrent connections. The results of the simulation were figured for all the nine *GT* output parameters, and were compared to both the *Train* and *Test* datasets. The goal was to see how accurately the *RNN* outputs follow both the *Train* and *Test* datasets trends. According to the results, the optimal *RNN* model was achieved after 1200 iterations, with eight neurons in the hidden layer, a recurrent connection with delay of 1 time-step in the hidden layer, and a recurrent connection with delay of 1 and 2 time-steps from the output to the first layer.

The *root mean squared relative error (RMSRE)* of the results for the *RNN* model was calculated according to Equation (1), where n_d is the number of datasets, y is the available (simulated) data, and y_m is the prediction of the *RNN* model.

$$RMSRE\ (\%) = 100 * \sqrt{\frac{1}{n_d} \sum_{i=1}^{n_d} \left(\frac{y - y_m}{y}\right)^2} \qquad (1)$$

5 Results

Figures 6 to 14 compare outputs of the *GT* with the outputs of *RNN 4-H-9*, for both the *Train* and *Test* datasets (targets). Figures 6 to 8 show the comparison between outputs of the *RNN* model and the *GT Train* and *Test* data for temperatures at different operational points of the *GT* (T_{02}, T_{03}, T_{04}). In Figures 9 to 11, the results are shown for the pressure variables (p_{02}, p_{03}, p_{04}). Figures 12, 13, and 14 indicate the same comparison for the *compressor pressure ratio (CPR)*, *rotational speed (N)*, and *GT efficiency (E_{GT})* respectively. Comparisons between outputs of the *GT* and the *RNN* model for the *Train* datasets demonstrate satisfactory results for the training process. The resulting *RNN* model was validated against the *Test* datasets. According to Figures 6 to 14, the validation results are also satisfactory for all *GT* output parameters. As these figures show, the outputs of *RNN* 4-H-9 follow the targets very closely for both the *Train* and *Test* datasets.

Figure 15 shows a comparison between *RMSRE (%)* of *RNN 4-H-9* outputs for the *Train* and *Test* datasets. According to this figure, *the average RMSRE (%) of the nine output variables (Ave.) for the Train and Test* datasets are about 0.22% and 2.6% respectively. It demonstrates that the *RNN* model developed in this study has a high reliability and accuracy in capturing the system dynamics.

6 Conclusions

In this study, a recurrent neural network of a single shaft gas turbine was investigated and developed in *Python* programming environment by using *Pyrenn* Toolbox, and employing 3000 simulated datasets for thirteen significant variables of the *GT*. The resulting model was validated against the *Test* datasets. The results demonstrated that response of the *RNN* model to variations in input parameters followed the system outputs with an acceptable accuracy. It proves that the *RNN* model is capable of performance prediction of the system with a high reliability. The methodology employed in this study provides a simple and reliable approach in dynamic modelling and simulation of gas turbines, especially when access to operational datasets is limited for any reason.
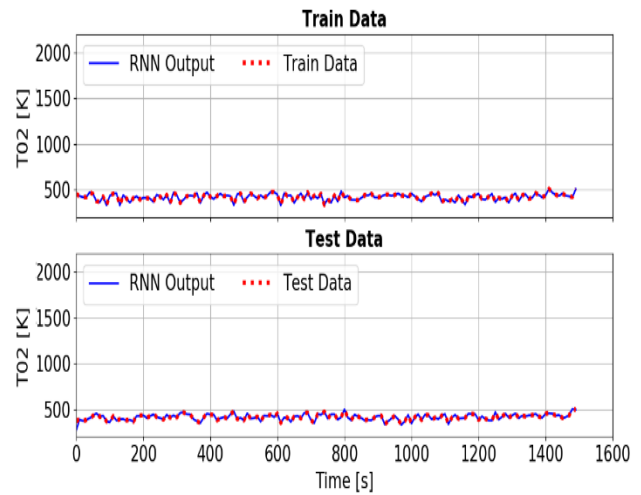
Figure 6. A comparison between output of the *RNN* model and the *GT Train* and *Test* datasets for T_{02}.
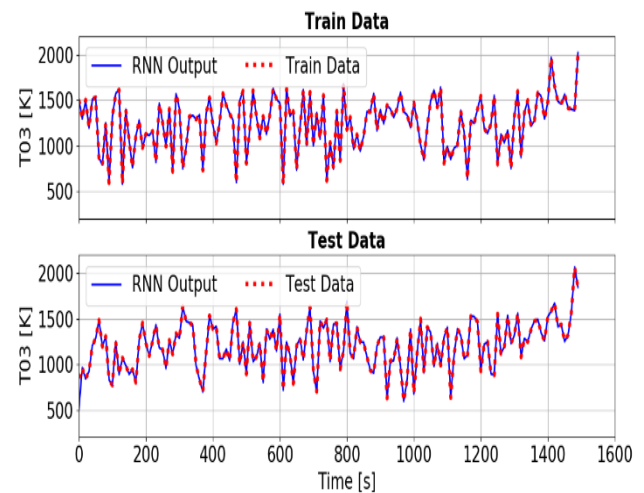
Figure 7. A comparison between output of the *RNN* model and the *GT Train* and *Test* datasets for T_{03}.
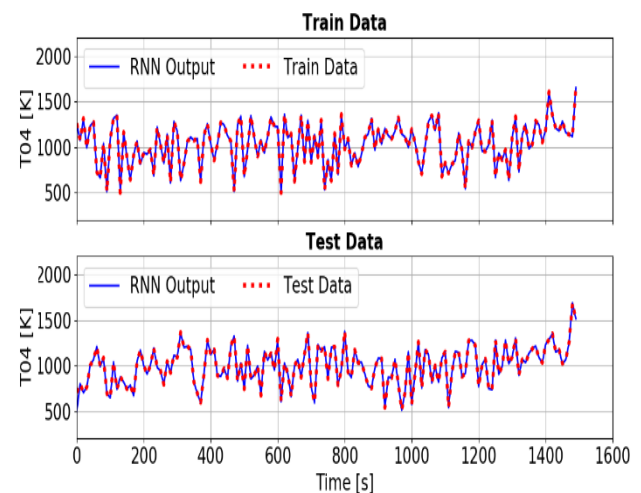
Figure 8. A comparison between output of the *RNN* model and the *GT Train* and *Test* datasets for T_{04}.
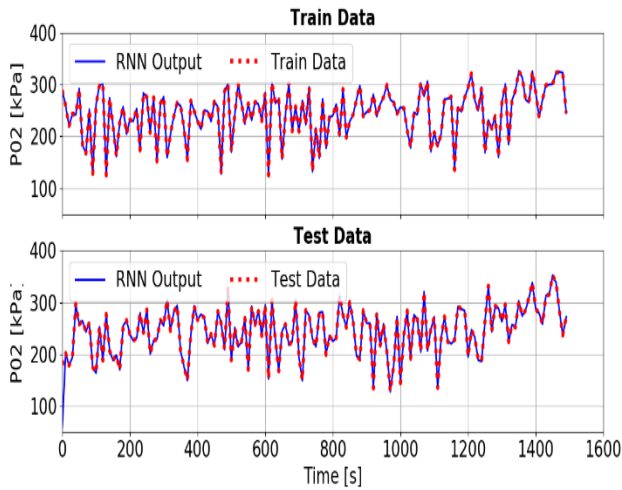
Figure 9. A comparison between output of the *RNN* model and the *GT Train* and *Test* datasets for P_{02}.

Figure 10. A comparison between output of the *RNN* model and the *GT Train* and *Test* datasets for P_{03}.

Figure 11. A comparison between output of the *RNN* model and the *GT Train* and *Test* datasets for P_{04}.

Figure 12. A comparison between output of the *RNN* model and the *GT Train* and *Test* datasets for compressor pressure ratio (*CPR*).

Figure 13. A comparison between output of the *RNN* model and the *GT Train* and *Test* datasets for rotational speed (*N*).

Figure 14. A comparison between output of the *RNN* model and the *GT Train* and *Test* datasets for *GT* efficiency (E_{GT}).

Figure 15. *RMSRE (%)* of *RNN 4-H-9* outputs for the *GT Train* and *Test* datasets.

References

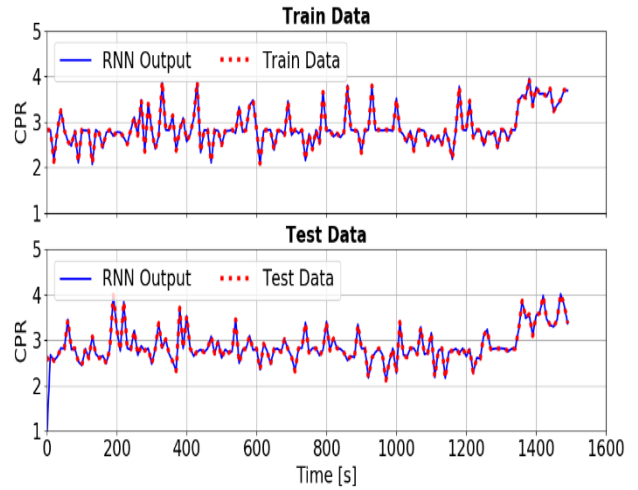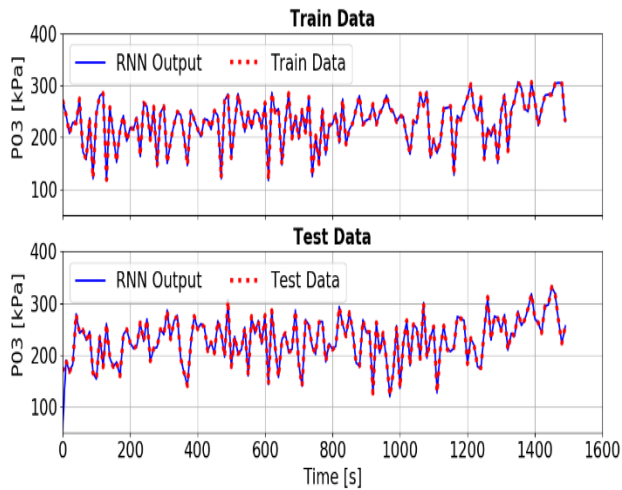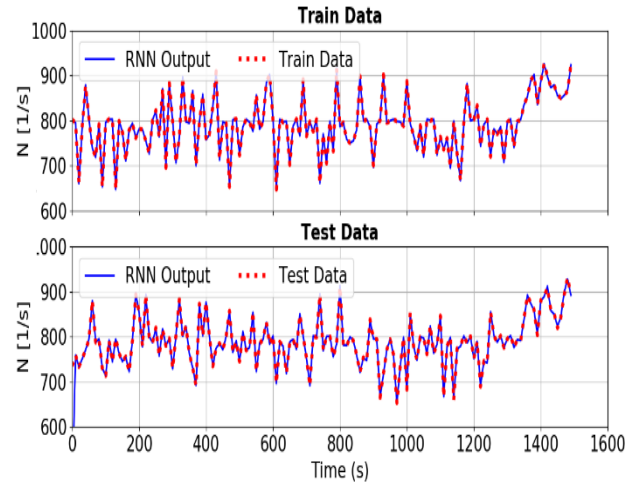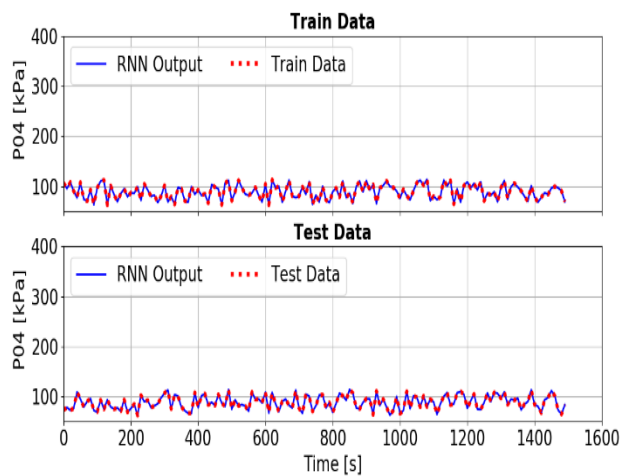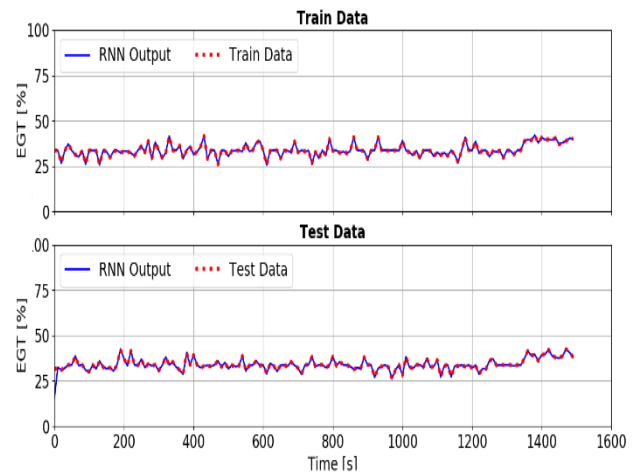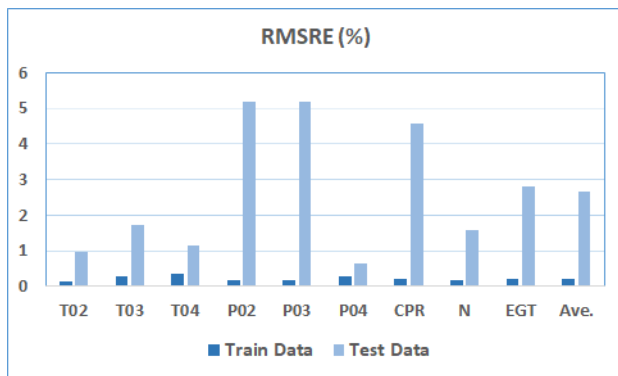R. K. Agrawal and M. Yunis. A Generalized Mathematical Model to Estimate Gas Turbine Starting Characteristics. *ASME Journal of Engineering for Power*, 104(1):194-201, 1982. doi:10.1115/1.3227249.

P. Ailer, I. Santa, G. Szederkenyi and K. M. Hangos. Nonlinear Model-Building of a Low-Power Gas Turbine. *Periodica Polytechnica Transportation Engineering*, 29(1-2):117-135, 2002. URL:eprints.sztaki.hu/2872.

S. M. Arabi, H. Ghadamian, M. Aminy, H. A. Ozgoli, B. Ahmadi and M. Khodsiani. The Energy Analysis of GE-F5 Gas Turbines Inlet Air-Cooling Systems by the Off-Design Method. *Measurement and Control*, 52(9-10):1489-1498, 2019. doi:10.1177/0020294019877485.

J. Arriagada, M. Genrup, A. Loberg and M. Assadi. Fault Diagnosis System for an Industrial Gas Turbine by Means of Neural Networks. *Proceedings of the International Gas Turbine Congress*, Tokyo, Japan, 2003.

H. Asgari, X. Q. Chen, M. B. Menhaj and R. Sainudiin. ANN-based System Identification, Modelling and Control of Gas Turbines - A Review. *Advanced Materials Research*, Vol. 622-623, pp. 611-617, 2012. doi:10.4028/www.scientific.net/amr.622-623.611.

H. Asgari, X. Q. Chen and R. Sainudiin. Modelling and Simulation of Gas Turbines. *International Journal of Modelling, Identification, and Control (IJMIC)*, 20(3):253-270, 2013a. doi:10.1504/ijmic.2013.057137.

H. Asgari, X. Q. Chen, M. B. Menhaj and R. Sainudiin. Artificial Neural Network Based System Identification for a Single-Shaft Gas Turbine. *ASME Journal of Engineering for Gas Turbines and Power,* 135(9):092601 (7 pages), 2013b. doi:10.1115/1.4024735.

H. Asgari, X. Q. Chen, R. Sainudiin, M. Morini, M. Pinelli, P. R. Spina and M. Venturini. Modelling and Simulation of the Start-Up Operation of a Heavy-Duty Gas Turbine Using NARX Models. *Proceedings of ASME Turbo Expo 2014*, Vol. 3A, GT2014-25056, 10 pages, Düsseldorf, Germany, 2014. doi:10.1115/gt2014-25056.

H. Asgari. Modelling, Simulation, and Control of Gas Turbines Using Artificial Neural Networks. *PhD Thesis*, University of Canterbury, Christchurch, New Zealand, 230 pages, 2014. URL:http://hdl.handle.net/10092/9355.

H. Asgari, M. Venturini, X. Q. Chen and R. Sainudiin. Modeling and Simulation of the Transient Behavior of an Industrial Power Plant Gas Turbine. *ASME Journal of Engineering for Gas Turbines and Power*, 136(6):061601 (10 pages), 2014. doi:10.1115/1.4026215.

H. Asgari and X. Q. Chen. Gas Turbines Modelling, Simulation, and Control: Using Artificial Neural Networks. *Taylor & Francis*, CRC Press, USA, 206 pages, ISBN 9781138893443, 2015. doi:10.1201/b18956.

H. Asgari, X. Q. Chen, M. Morini, M. Pinelli, R. Sainudiin, P. R. Spina and M. Venturini. NARX Models for Simulation of the Start-Up Operation of a Single-Shaft Gas Turbine. *Applied Thermal Engineering*, Vol. 93, pp. 368-376, 2016. doi:10.1016/j.applthermaleng.2015.09.074.

H. Asgari, M. Fathi and X. Q. Chen. Design of Conventional and Neural Network Based Controllers for a Single-Shaft Gas Turbine. *Aircraft Engineering and Aerospace Technology*, Emerald Publishing Limited, 89(1):52-65, 2017. doi:10.1108/aeat-11-2014-0187.

M. R. B. Tavakoli, B. Vahidi and W. Gawlik. An Educational Guide to Extract the Parameters of Heavy-Duty Gas Turbines Model in Dynamic Studies Based on Operational Data. *IEEE Transactions on Power Systems*, 4(3):1366-1374, 2009. doi:10.1109/tpwrs.2009.2021231.

C. M. Bartolini, F. Caresana, G. Comodi, L. Pelagalli, M. Renzi and S. Vagni. Application of Artificial Neural Networks to Micro Gas Turbines. *Energy Conversion and Management,* 52(1):781-788, 2011. doi:10.1016/j.enconman.2010.08.003

M. Basso, L. Giarre, S. Groppi and G. Zappa. NARX Models of an Industrial Power Plant Gas Turbine. *IEEE Transactions on Control Systems Technology*, 13(4):599-604, 2004. doi:10.1109/tcst.2004.843129.

M. B. Rahmoune, A. Hafaifa and G. Mouloud. Fault Diagnosis in Gas Turbine Based on Neural Networks: Vibrations Speed Application. *Advances in Acoustics and Vibration*, Springer, Vol. 5, pp. 1-11, 2017. doi:10.1007/978-3-319-41459-1_1.

N. Chiras, C. Evans and D. Rees. Nonlinear Modelling and Validation of an Aircraft Gas Turbine Engine. *Nonlinear Control Systems*, 34(6):871-876, IFAC Symposia Series, 2001. doi:10.1016/s1474-6670(17)35289-8.

N. Chiras, C. Evans and D. Rees. Nonlinear Gas Turbine Modelling Using Feedforward Neural Networks. *Proceedings of ASME Turbo Expo 2002*, Vol. 2, GT2002-30035, pp. 145-152, Amsterdam, Netherlands, 2002. doi:10.1115/gt2002-30035.

N. Chiras, C. Evans and D. Rees. Global Nonlinear Modelling of Gas Turbine Dynamics Using NARMAX Structures. *ASME Journal of Engineering for Gas Turbines and Power*, 124(4):817-826, 2002. doi:10.1115/1.1470483.

S. Cisottoy and R. Herzallah. Performance Prediction Using Neural Network and Confidence Intervals: A Gas Turbine Application. *Proceedings of 2018 IEEE International Conference on Big Data*, pp. 2151-2159, Seattle, WA, USA, 2018. doi:10.1109/BigData.2018.8621919.

W. H. Elashmawi, N. A. Kotp and G. E. Tawel. Neural Network Monitoring Model for Industrial Gas Turbine. *Asian Journal of Applied Sciences*, 5(3):608-614, 2017. URL:http://orcid.org/0000-0003-0142-0632.

W. H. Elashmawi. A Novel Proposed Neural Network MAD (Monitoring, Analysis and Diagnose) Model for Industrial Gas Turbine. *International Journal of Soft Computing*, 13(3):92-101, 2018. doi:10.36478/ijscomp.2018.92.101.

Encyclopedia Britannica Inc., 1999. Retrieved from: *www.britannica.com/technology/gas-turbine-engine*.

M. Fast, M. Assadi and S. De. Condition Based Maintenance of Gas Turbines Using Simulation Data and Artificial Neural Network: A Demonstration of Feasibility. *Proceedings of ASME Turbo Expo 2008*, GT2008-50768, pp. 153-161, Berlin, Germany, 2008. doi:10.1115/gt2008-50768.

M. Fast, M Assadi and S. De. Development and Multi-Utility of an ANN Model for an Industrial Gas Turbine. *Applied Energy*, 86(1):9-17, 2009. doi:10.1016/j.apenergy.2008.03.018

M. Fast and T. Palmé. Application of Artificial Neural Network to the Condition Monitoring and Diagnosis of a Combined Heat and Power Plant. *Energy*, 35(2):1114-1120, 2010. doi:10.1016/j.energy.2009.06.005.

T. Palmé, M. Fast and A. Karlsson. Gas Turbines Sensor Validation through Classification with Artificial Neural Networks, *Applied Energy*, 88(11):3898-3904, 2011. doi:10.1016/j.apenergy.2011.03.047.

Fortum, and VTT Technical Research Centre of Finland Ltd. 2020. Retrieved from: *www.apros.fi/en/*.

M. A. Ibrahem, O. Akhrif, H. Moustapha and M. Staniszewski. Neural Networks Modelling of Aero-derivative Gas Turbine Engine: A Comparison Study. *Proceedings of the 16th International Conference on Informatics in Control, Automation and Robotics*, Vol. 1: ICINCO, pp. 738-745, Prague, Czech Republic, 2019. doi:10.5220/0007928907380745.

F. Jurado. Nonlinear Modelling of Micro-Turbines Using NARX Structures on the Distribution Feeder. *Energy Conversion and Management*, 46(3):385-401, 2005. doi:10.1016/j.enconman.2004.03.012.

S. Kiakojoori and K. Khorasani. Dynamic Neural Networks for Gas Turbine Engine Degradation Prediction. Health Monitoring and Prognosis. *Neural Computing & Applications*, 27(8):2157–2192, 2016. doi:10.1007/s00521-015-1990-0.

J. H. Kim, T. W. Song, T. Kim and S. T. Ro. Model Development and Simulation of Transient Behavior of Heavy Duty Gas Turbines. *ASME Journal of Engineering for Gas Turbines and Power,* 123(3):589-594, 2001. doi:10.1115/1.1370973.

J. H. Kim, T. W. Song, T. Kim and S. T. Ro. Dynamic Simulation of Full Startup Procedure of Heavy-Duty Gas Turbines. *ASME Journal of Engineering for Gas Turbines and Power*, 124(3):510-516, 2002, doi: 10.1115/1.1473150.

A. Lazzaretto and A. Toffolo. Analytical and Neural Network Models for Gas Turbine Design and Off-Design Simulation. *International Journal of Thermodynamics*, 4(4):173-182, 2001.

Z. Liu and I. A. Karimi. Gas Turbine Performance Prediction via Machine Learning. *Energy*, Vol. 192, No. 116627 (10 pages), 2020. doi:10.1016/j.energy.2019.116627.

S. O. T. Ogaji, R. Singh and S. D. Probert. Multiple-Sensor Fault-Diagnosis for a 2-Shaft Stationary Gas Turbine. *Applied Energy*, 71(4):321-339, 2002. doi: 10.1016/S0306-2619(02)00015-6.

A. E. Ruano, P. J. Fleming, C. Teixeiraa, K. R. Vazquezc and C. M. Fonsecaa. Nonlinear Identification of Aircraft Gas Turbine Dynamics. *Neurocomputing*, 55(3–4):551-579, 2003, doi:10.1016/S0925-2312(03)00393-x.

A. Salehi and M. Montazeri. Black Box Modeling of a Turboshaft Gas Turbine Engine Fuel Control Unit Based on Neural NARX. *Journal of Engineering for the Maritime Environment*, 233(3):949-956, 2018. doi:10.1177/1475090218797496.

S. Simani and R. Patton. Fault Diagnosis of an Industrial Gas Turbine Prototype Using a System Identification Approach. *Control Engineering Practice*, 16(7):769-786, 2008. doi:10.1016/j.conengprac.2007.08.009.

Soumalya Sarkar, K. Mukherjee, Soumik Sarkar and A. Ray. Symbolic Transient Time-Series Analysis for Fault Detection in Aircraft Gas Turbine Engines. *Proceedings of American Control Conference (ACC)*, pp. 5132-5137, Montreal, QC, Canada, 2012. doi:10.1109/acc.2012.6315253.

Soumalya Sarkar, K. Mukherjee, Soumik Sarkar and A. Ray. Symbolic Dynamic Analysis of Transient Time-Series for Fault Detection in Gas Turbine Engines. *Journal of Dynamic Systems, Measurement, and Control*, 135(1): 014506 (6 pages), 2013. doi:10.1115/1.4007699.

K. Suzuki. Artificial Neural Networks - Industrial and Control Engineering Applications. *IntechOpen*, Croatia, 492 pp., ISBN: 978-953-307-220-3, 2011.

Y. Tiumentsev and M. Egorchev. Neural Network Modeling and Identification of Dynamical Systems. *Academic Press*, 332 pp., ISBN: 9780128154304, 2019.

G. Torella, F. Gamma and G. Palmesano. Neural Networks for the Study of Gas Turbine Engines Air System. *Proceedings of the International Gas Turbine Congress*, 8 pages, Tokyo, Japan, 2003.

P. P. Walsh and P. Fletcher. Gas Turbine Performance. *Blackwell Science*, Osney Mead, Oxford OX2 0EL, UK, 628 pp., 1998. doi:10.1017/S0001924000027834.

Y. Yoru, T. H. Karakoc and A. Hepbasli. Application of Artificial Neural Network (ANN) Method to Exergetic Analyses of Gas Turbines. *International Symposium on Heat Transfer in Gas Turbine Systems*, 4 pages, Antalya, Turkey, 2009. doi:10.1615/ICHMT.2009.HeatTransfGasTurbSyst.580.

Simulation study for comparison of control structures for BFB biomass boiler

Milan Zlatkovikj Valentina Zaccaria Ioanna Aslanidou Konstantinos Kyprianidis

School of Business, Society and Engineering, Malardalen University, Vasteras, Sweden
`{milan.zlatkovikj, valentina.zaccaria, ioanna.aslanidou,`
`Konstantinos.kyprianidis}@mdh.se`

Abstract

Biomass fired boilers usage is increasing due to supportive policies and economic trends. Fluidized bed technology is identified as proper solution for lower quality fuels such as biomass. Moisture and heating value can vary significantly in biomass fuels. Without real-time information on their variation, they are a disturbance to the system. These disturbances affect the system steady state and decrease operational efficiency. Proper characterization of the disturbance enables the use of Feed-Forward control. Feed-Forward makes use of the knowledge about the updated condition of the fuel and can act towards reducing the impact of the fuel on offsetting the system. Feed-Forward Model Predictive Control is proposed as new control strategy. Comparison is made between the existing control strategy and the new proposed solution. Control performance is evaluated on three process outputs, in three different scenarios. Adding feed-forward signal for fuel moisture improves control performance in both controllers, while ultimately Feed-Forward Model Predictive Control shows the best performance in most comparison metrics.

Keywords: biomass fuel, fuel moisture, model predictive control, feed-forward, plant control

1 Introduction

Biomass usage as a fuel is constantly growing due to supportive government policies and positive market trends. The highest share of biomass utilization as fuel goes in combustion in large scale steam boilers, used in Combined Heat and Power (CHP) plants (Atsonios et al., 2020).

Fluidization technology has been implemented towards creating more favorable combustion conditions in the furnace (Leckner, 2003). Bubbling Fluidized Bed (BFB) and Circulating Fluidized Bed (CFB) boilers operate on this technology. CFB are the newer, improved version of the two, and therefore are more applied and more work is done on their analysis. However, for applications with lower power output, BFB boilers are still applicable and competitive (Peña, 2011). The differences between BFB and CFB boilers are in the power output, geometry of furnace, layout (configuration of heat exchangers), values of operational parameters, and there are differences in flow streams (mostly noted in the recirculation part) (Peña, 2011). Except for the main operational differences, there are noted differences in the research work done, such as analyzing the capability of power output change over time, temperature profiles in the furnace, and thermal capacity of the systems (Huttunen et al., 2017; Arena et al.,1995).

Process modeling is a cost-effective way to analyze system performance, compared to real tests at operating large scale facilities. The key requirement before process modeling is to specify the aim of the modeling work, since there are plenty of different types of models, and just the right combination of model and application can help towards effective process analysis (Atsonios et al., 2020). Simplified mass and energy balances models, with 0 dimensionality (e.g. divided in several control volumes of interest), are a simple and effective way for boiler system analysis (Sandberg et al., 2011). Although they provide less details compared to CFD and other high dimensional models that include more process features, their simulation time is shorter than CFD models and they can capture the process behaviors and be utilized e.g. for control or optimization purpose. Models that can be simulated much faster than real-time can be implemented on-line, as a prediction support for the controller (Szentannai, 2011; Zimmerman et al., 2018).

Most of the applied controllers in thermal power plant are based on Proportional Integral (PI) algorithm (Szentannai, 2011). This type of control is proven in operation and reliable. However, with the new demands for the power plants flexibility due to the rise of use of intermittent renewable energy sources, the current control structure efficiency is limited. PI controllers are known for their limitations for multi-input multi-output (MIMO) processes with strong coupling between variables, such as steam boilers (Szentannai, 2011).

On the other hand, Model Predictive Control (MPC) is perceived as a good option for replacement of the current controllers, improving control performance in strongly coupled systems (Qin and Badgewell, 2013). They are proven in applications in the chemistry industry, while their application in power plants is lagging behind (Szentannai, 2015). One of the main

identified difficulties towards implementation of MPC is the deficit of quality dynamic models, on which the controller is based (Atsonios et al., 2020). In MPC, the controller relies on the accuracy of the underlying model. In particular, the performance can degrade in presence of unmodeled disturbances. Fuel quality (heating value and moisture) is the strongest disturbance in biomass combustion systems.

The necessity for increase of research work on dynamic modeling of biomass energy applications has been emphasized in a recent review (Atsonios et al., 2020). The amount of work published on this topic is scarce compared both to dynamic studies on fossil fuel systems and steady operation studies on biomass. In addition to this, many research groups develop separate models on sub-systems of the boiler, which does not allow for full system analysis and communication between different types of models used.

Combustion analysis on BFB boilers with biomass and waste fuels have been analyzed in (Ravelli et al., 2008; Galgano et al., 2005; Scala and Chirone, 2004). These models provide detailed analysis of the combustion process. However, they are too computationally demanding for implementation in control purposes.

Control work focusing on the water/steam side dynamics has been studied in (Åström and Bell, 2000). In this work, the combustion part of the boiler is simplified.

As mentioned above, modeling the disturbance to the process can enhance the control. The major uncertainty in the boiler is related to the heat input, which comes from fuel moisture variations. By having a model that includes information about the moisture variation in the incoming fuel, we have additional important parameter knowledge, which can help toward a better analysis of the system. The benefits of adding feed-forward signal to MPC controller are analyzed in (Carrasco and Goodwin, 2011). Combining the knowledge of the fuel moisture with advanced control method can help towards decreasing the effect of fuel variation on the process outputs.

The aim of this work is therefore to show how MPC can be applied in biomass-fueled BFB boilers, how it compares to conventional PI control, and quantify the benefit introduced by feed-forward (FF). In the following sections, the analyzed process and its characteristics are described, an overview of the numerical model and the control structures are given, after which results are presented and discussed.

2 Process description

The analyzed system is a Bubbling Fluidized Bed (BFB) Boiler fired by mix of woody fuels (bark, chips). It is used in Eskilstuna Strängnäs Energi och Miljö AB (ESEM), CHP plant, to provide electricity and district heating for the local municipality.

Its nominal operation parameters are listed in Table 1. Throughout the year the plant operates with change of the load (boiler power output) based on demand and plant condition. The boiler layout is presented in Figure 1. The boiler consists of a fluidized bed region, freeboard, a superheaters section (SH 1,2,3) and an cconomizer.

Table 1. Boiler nominal parameters

Parameters	Value
Boiler power output	110 MW
Final Steam Temperature	540 ºC
Fuel Flow	13.7 kg/s
Steam flow	41.1 kg/s
Fuel LHV, wet basis	8500 kJ/kg
Fuel moisture	48%

The fuel is delivered into the fluidized bed region. In BFB boilers, a staged combustion is applied, where air is delivered at three different heights – primary, secondary and tertiary air. In addition, there is recirculation of flue gasses, which is used for temperature regulation in the fluidized bed region. Combustion process is completed along the freeboard, and after that, the hot flue gasses transfer the heat to the water/steam side, at the steam superheaters and the economizer.
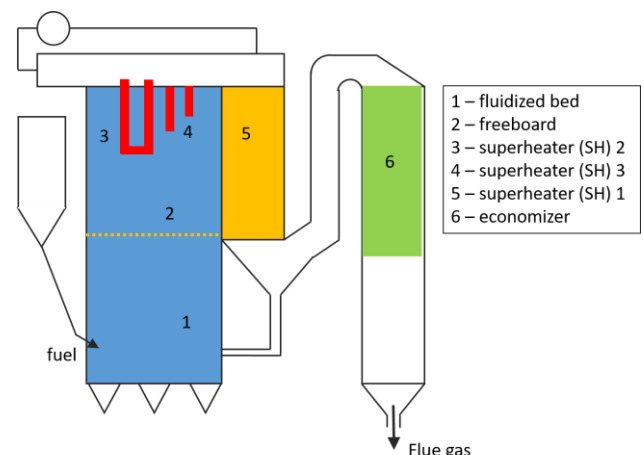
Figure 1. Boiler Layout

Water is pumped from the water tank, goes for preheating in the economizer, evaporates to steam in the evaporator pipes placed around the furnace and finally is superheated in 3 superheaters before entering the steam turbine. There are 2 steam attemperators (water sprays injected in the steam), to control steam temperature.

3 Methodology

3.1 Model description

Developing models for big industrial plants, such as steam boilers is a time demanding task. Given that

boilers are extensively used for various applications and the effort to develop models of these plants, it is important to make sure of the model's reusability (Casella et al., 2014). Models developed with acausal approach allow for reusability with less effort compared to making a new one from scratch (Casella et al., 2014).

The programming language Modelica and the software Dymola are used in this work for that aim. With minor adjustments, the model can be used to simulate different system types and configurations. In addition, the model can be exported to various other software, to develop control or other frameworks. In this work, the Dymola model is exported to Simulink, where system identification (SI) and controller design are performed.

3.2 Boiler equations

Based on the Boiler layout, the system is divided in control volumes to describe the process characteristics in a simplified way and to capture all essential aspects of the process.
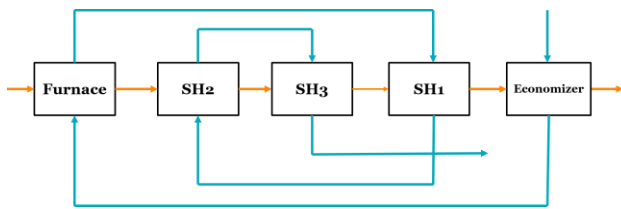

Figure 2. Boiler flows logics

The logic of the flows analyzed within the model is shown in Figure 2. Orange lines represent the path of the hot side (flue gas) and the blue lines represent the cold side (water/steam).

$$\Delta T = \frac{(T_{h1} - T_{c2}) - (T_{h2} - T_{c1})}{\ln\left(\frac{T_{h1} - T_{c2}}{T_{h2} - T_{c1}}\right)} \ [K] \tag{1}$$

$$\frac{dT}{dt} = \frac{Q_{balance}}{HE_{capacity}} = \frac{Q_{in} - Q_{out}}{cp * m}$$
$$= \frac{Q_{in} - Q_{out}}{cp * (\rho * V)} \left[\frac{K}{s}\right] \tag{2}$$

$$Q = A * U * \Delta T \ [MW] \tag{3}$$

The temperature difference in the heat exchangers is calculated based on equation (1), where h is used for the hot fluid (flue gasses) and c for the cold fluid (water and steam). Subscripts 1 and 2 indicate inlet and outlet value for the parameter respectively. Q is used for exchanged heat, A is surface of heat exchanger, U is overall heat transfer coefficient and T is temperature. Using equations (1-3), energy balances of all components of interest are calculated.

The key part of boiler analysis is fuel analysis. This boiler uses a mix of woody fuels.

In regular operation, moisture is analyzed for all incoming shipments of wood to the plant by taking random samples and measuring the moisture content. With this approach, the plant has documented the yearly variation of the moisture, on average per month. Yearly variations of moisture are displayed in Figure 3. Total heat input in the furnace is obtained as a product of lower heating value (LHV) of fuel, calculated using the moisture content reading and the fuel mass flow.
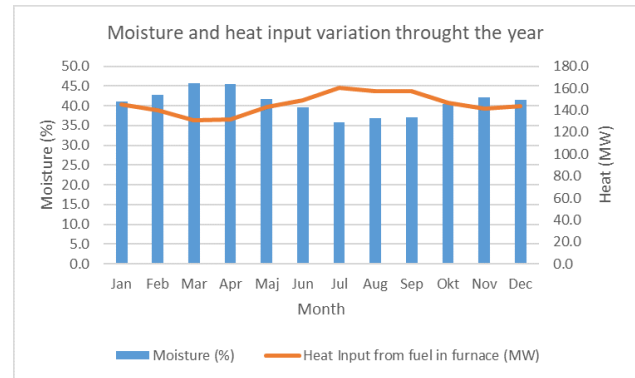
Figure 3. Yearly variation of moisture and heat input

However, this information doesn't allow for analysis with 1min sampling time. The moisture content has the highest variation in woody fuels, while the other components of elemental analysis (C, H, O, N, S and ash) have minimal variation.

In this work, results from lab analysis of the fuel used in the plant, are used for values of elemental analysis components on dry basis. The moisture content value can be read from the spectra sensor installed on-line above the conveyor belt. Using the moisture value and fuel conversion equations, we obtain values for fuel composition on wet basis (as received) and use the values to calculate the fuel heating value, hence simulating moisture variations in real-time.

The fuel heating values (LHV and HHV) are calculated by (Saidur et al., 2011):

$$HHV = 0.3491 * X_C + 1.1783X_H \\ + 0.1005X_S - 0.0151 * X_N \\ - 0.1034 * X_O - 0.0211 \\ * X_{ash} \ [MJ/kg] \tag{4}$$

$$LHV = HHV * \left(1 - \frac{w}{100}\right) - 2.444 * \frac{w}{100} \\ - 2.444 * \frac{h}{100} * 8.936 \\ * \left(1 - \frac{w}{100}\right) \left[\frac{MJ}{kg}\right] \tag{5}$$

where: X_i is wt.%. content of the fuel elements on dry basis, w is moisture content, and h is hydrogen content.

Enthalpies and specific heat capacities of the streams are calculated using polynomial functions (Wester, 2015). Other required parameters for the model are obtained from plant documentation (geometry of boiler and capacity of heat exchangers) and tuning parameters are tuned based on historical operational values. The model operates on simplification for full conversion of combustible fuel elements to flue gasses.

In order to capture the dynamics of the process and to have a suitable framework for comparison of the model, key inputs and outputs are identified. These are used during model validation, system identification and control simulations. Key parameters are listed in Table 2. Several key process indicators cannot be obtained from the plant because there are no present sensors for them. There is no measurement signal on fuel mass flow, temperature in the freeboard and mass flow of water sprays in the steam attemperators. Fuel mass flow is estimated based on other related parameters in the energy balance. Due to the lack of measured signals, some common comparisons for boiler performance, such as temperature profile in the furnace are not feasible. Selected outputs for comparison are chosen based on the available sensors: final steam temperature, flue gasses temperature (after superheaters stage), fluidized bed temperature and power output.

Table 2. Key boiler parameters for model

Inputs	Outputs
Fuel mass flow	Power output
Water mass flow	Steam final temp.
Fuel moisture (%)	Fluidized bed temp.
Air mass flow	Flue gas temp.
Recirc. Flue gas flow	

The final steam temperature is a key parameter for boiler operation. It is predefined from the manufacturer during construction and it can have small variations from nominal set-point, to ensure safe and efficient operation

Since there is no temperature measurement of the freeboard temperature, the flue gasses temperature after the superheaters section is the first indicator that can be used to assess the operation. This parameter combined with the steam parameters measurements allows to back-calculate the temperatures of the flue gasses at various points.

The fluidized bed temperature is the key parameter that indicates the operation stability for BFB boilers. This parameter has minimal variation, due to its characteristics – dense region with very high thermal capacity.

The power output is key for the control action. Most of the time, the power output is adjusted to satisfy supply and demand for electricity and heat, and

controlled change of setpoint is made to achieve it. It is calculated the balance betweenas the energy contained in the steam at the final point (after the last superheater, before turbine inlet) and the starting condition (feedwater tank).

After developing dynamic model for the analyzed plant, the model is tuned and adjusted based on the historical values of the operational parameters. Model validation is done by running developed model with operation data for the key inputs and comparing the output of the model against the operation data. In order to make use of the developed dynamic model for control, it needs to be converted in a suitable form, so that a control structure can work with it. For this work we use linear MPC, developed in Matlab, based on identified linear state-space model.

3.3 System identification

State space model is a mathematical form of a system that contains input, output and state variables. Equation 6 shows the form of Discrete-time identified state-space model, identified with n4sid function in Matlab:

$$\begin{cases} \dot{x} = Ax + Bu + Ke \\ y = Cx + Du + e \end{cases} \quad (6)$$

where: x is state vector, y is output vector, u is input vector, e is disturbance vector, A, B, C, D, and K are state-space matrices.

Subspace identification is used to identify linear state-space, which will be the basis for the MPC design. Important for the identified model is to have good fit with the output data and to capture process dynamics.

The inputs for the SI process were constituted by step changes in the chosen input signals, with increase and decrease of predefined ratios of 1, 2.5, 5 and 10%. Step signals are combined in a way to capture the impact of coupling inputs on the observed outputs. The input sequence generated for this work contains all combinations of inputs excitation. Inputs and outputs are used from Table 2, only fuel moisture is used as disturbance.

For the defined set of inputs and outputs, a 5th order state space model is identified with n4sid function in Matlab, and it represents the internal model for MPC design and tuning.

3.4 Control structures

Model predictive control (MPC) can be defined as class of control algorithms that utilizes an explicit process model to predict the future response of a plant (Qin and Badgewell, 2013). MPC uses a dynamic model of the plant to predict future actions of the manipulated variables on the plant output. MPC is defined with: process model, set of constraints and objective function. The function that describes the MPC can be found in literature under the names of objective, target or cost

function. This function provides information about the importance given to control error minimization and frequency of actuators operation. The process model should capture the dynamics of inputs, outputs and disturbances in the controlled process. The drawback is that MPC is more complex to develop compared to conventional PI control, and it requires more time to develop and tune properly.

PI (Proportional and Integral) Control is a commonly applied control structure in industry. It combines the benefits of proportional and integral control into one. PI scheme is presented in Figure 4. The PI controller does not receive signal about the disturbance in the system.
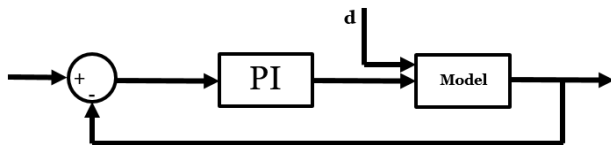

Figure 4. PI Control Scheme

A commonly used equation for PI controller is:

$$p(t) = \bar{p} + K_c \left(e(t) + \frac{1}{\tau_I} \int_0^t e(t^*) \, dt^* \right) \qquad (7)$$

where p(t) is the controller output, K_c is the proportional gain, e(t) is the measured error between output signal and setpoint, τ_I is the integral time.

The three PI controllers were tuned using an internal model control method and the control parameters were successively fine-tuned to improve the performance (Tan et al., 2006). For a full comparison, the PIs were integrated with feed-forward (FF) for disturbance rejection for case 2. The FF model was set as a first order input-output model according to equation 6.

$$p_{FF} = -d \cdot \frac{G_d}{G_p} \qquad (8)$$

Where p_{FF} is added to the PI output, d is the disturbance signal, and Gd and Gp are first-order transfer functions between d and y (process output), and p and y respectively. FF PI scheme is presented in Figure 5. The disturbance signal is added to the PI controller as FF.

The control parameters used for the PI control in both configurations is provided in Table 3. This table also contains information about the connection between inputs and outputs in the PI logics i.e. which input is manipulated to control a specified output.
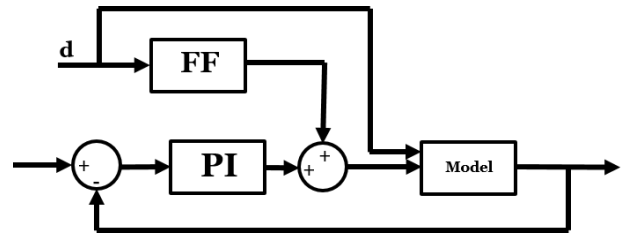
Figure 5. FF PI Control Scheme

The PI structure contains 3 SISO (Single Input Single Output) control loops, listed in Table 3.

Table 3. PI controller parameters

Control loop	Proportional K_c	Integral τ_I
Fuel mass flow Power Output	0.4 (-)	150 (s)
Water mass flow Final Steam temp.	-0.55 (-)	600 (s)
Total air mass flow Fluidized bed temp.	-0.5 (-)	1500 (s)

After importing the identified model in the MPC Toolbox in Matlab, its parameters are defined. Nominal values for the input and outputs are set, constraints on the inputs range and rate of change are set. Key MPC parameters used are provided in Table 4, while the constraints used for all controllers are listed in Table 5.

Table 4. MPC parameters

MPC Parameter	Value
Sampling time	60 (s)
Prediction horizon	10
Control horizon	2

The error between future outputs and target values is minimized by means of a quadratic objective function, presented in Equation 9:

$$J = W_j^{e_y} S_{e_y}(k)^2 + W_j^{MV} S_{MV}(k)^2 + W_j^{\Delta MV} S_{\Delta MV}(k)^2 \qquad (9)$$

The three terms in the objective function refer to the output error e_y (between controlled variables and targets), the manipulated variables (MV), and the change rate of manipulated variables. The weights Wj penalize each term in a different way.

Table 5. Constraints on manipulated parameters

Input	Nominal value (kg/s)	Control range constraints (kg/s)
Fuel flow	13.7	6 – 16
Water flow	35	20 - 40
Air flow	50	40 - 60

MPC control scheme is presented in Figure 6. The MPC controller does not receive signal about the measured disturbance in the system.

Figure 6. MPC Control Scheme

FF MPC control scheme is presented in Figure 7. The disturbance signal is added to the MPC controller as FF.
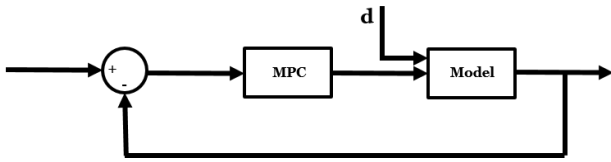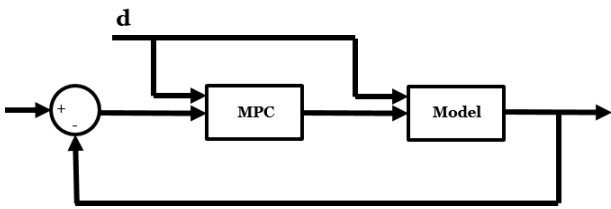

Figure 7. MPC FF Control Scheme

4 Results

After developing the control structures, they are compared in various scenarios, for their capability to deal with system disturbances and set points change. The moisture content (%) in the fuel is the system disturbance. Four types of controllers are compared – PI, FF PI, MPC and FF MPC. A list of scenarios, description of changes made in each setpoint and disturbance signal is provided in Table 6.

The controllers are analyzed in these scenarios for control and statistical metrics. Control metrics are overshoot, rise and settling times, which are common indices for evaluating controller performance, while statistical metrics used are the mean absolute error and standard deviation. Control metrics are chosen as standard control performance test, while statistical metrics can show us how much variation in controlled outputs we can expect during control scenarios and control actions. Since in scenario 3 we analyze the performance of controller to keep constant setpoint rather than response to step change, in this scenario we provide only statistical metrics.

The overshoot values for Scenarios 1 and 2 are displayed in Table 7. The values are expressed in % of the new set point and represent the height of the first peak after crossing the new setpoint value.

Rise time is defined as the time required for the output to change from 10% to 90% from the initial to the new setpoint.

Settling time is considered as the time required to reach 2% error between the signal and the new setpoint value.

Since settling and rise time are related to the change of setpoint, they are analyzed only for scenarios 1 and 2. Settling and rise times are shown in Tables 8 and 9 respectively. Due to limited space, statistical indicators and plots of inputs and outputs are shown only for Scenario 3. Absolute mean error is shown in Table 10, while standard deviation is in Table 11. Inputs are plotted in Figure 8, while outputs are plotted in Figure 9.

The simulations are done towards assessing the capability of the 4 controllers to deal with different control scenarios. Results are reported on 60s sampling time. It can be noted that results differ in the scenarios and the criteria for evaluation. This shows that we can't label one control option as absolutely best rather the best one for a specific application or in testing scenario.

It has to be noted that no weights are added in MPC tuning during this work to favorize one input or output on the expense of the rest. Adding weights on the inputs and the outputs would cause the controller to put more effort towards one parameter than the others and would make the comparison uneven.

Scenario 3 is the closest to real operation control actions. The disturbances are frequent during operation, and they come with different magnitudes due to the random uniform signal block used in Simulink. This signal is generated with sampling time of 50min. It can be said that Scenario 3 is solid basis towards more demanding control actions analysis.

Overshoots analysis shows that the lowest overshoot for all controllers is in output 3 – fluidized bed temperature, due to the huge thermal inertia of this region and slow changes.

On the settling time analysis, it takes the longest time to settle to the specified new setpoint, which is to be expected due to the same characteristic that makes the minimal or no overshoot.

From the results of scenario 3, we can observe that in statistical parameters, controllers with FF show better performance compared to the ones without FF – lower error and deviation. MPC FF has the lowest error and deviation in this Scenario. In all 3 scenarios, PI control has the worst control on output 3 – temperature of fluidized bed. Except for this parameter, we can see that for the other parameters the difference between MPC and PI controllers is not big.

Table 6. Analyzed Control scenarios

Scenario	Steam temp. setpoint [C]	Power setpoint [MW]	Fluid. Bed setpoint [C]	Moisture disturbance [%]
1	Step: 535 – 520	constant	constant	constant
	constant	Step: 110 – 100	constant	constant
	constant	constant	Step: 910 - 900	constant
2	Step: 535 – 520	constant	constant	Step: 48 – 53
	constant	Step: 110 – 100	constant	Step: 48 – 53
	constant	constant	Step: 910 - 900	Step: 48 – 53
3	constant	constant	constant	Random uniform input

Table 7. Overshoot (%)

Scenario	Scenario 1			Scenario 2		
Controller	Steam temp.	Power	FB. temp.	Steam temp.	Power	FB. temp.
1. PI	0	0.04	0	3.4	1	0.8
2. FF PI	0	0.04	0	1.1	0.04	0.7
3. MPC	0.17	0.2	0	0.82	2.6	0.4
4. FF MPC	0	0.3	0	0.16	0.67	0.1

Table 8. Settling time (s)

Scenario	Scenario 1			Scenario 2		
Controller	Steam temp.	Power	FB. temp.	Steam temp.	Power	FB. temp.
1. PI	720	1080	4620	1680	540	8640
2. FF PI	720	1080	4620	2340	840	8340
3. MPC	480	600	1680	3720	1680	4380
4. FF MPC	1320	540	960	1680	1380	3900

Table 9. Rise time (s)

Scenario	Scenario 1			Scenario 2		
Controller	Steam temp.	Power	FB. temp.	Steam temp.	Power	FB. temp.
1. PI	180	180	2220	120	120	300
2. FF PI	180	180	2160	180	180	360
3. MPC	420	120	180	120	120	120
4. FF MPC	720	240	660	180	180	420

Table 10. Mean absolute error (-) for Scenario 3

Controller	Steam temp.	Power output	Fluidized bed temp.
1. PI	0.0648	0.0523	0.7130
2. FF PI	0.1601	0.0196	0.6885
3. MPC	0.3566	0.0644	0.2237
4. FF MPC	0.0407	0.0079	0.0513

Table 11. Standard deviation (-) for Scenario 3

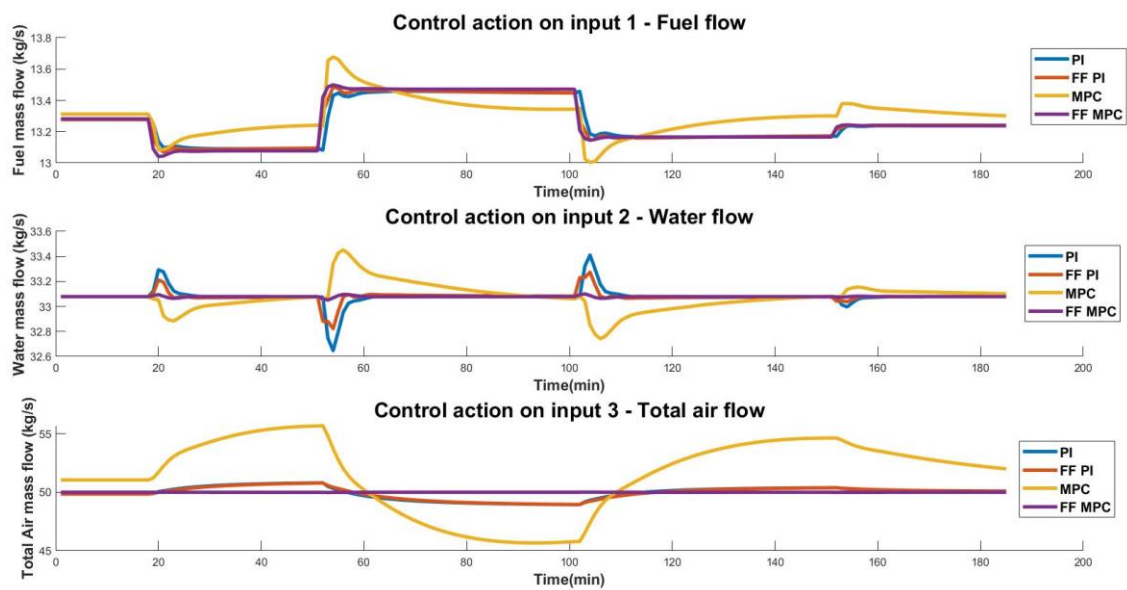| Controller	Steam temp.	Power output	Fluidized bed temp.
1. PI	0.1476	0.1449	0.8622
2. FF PI	0.2322	0.0584	0.8364
3. MPC	0.6253	0.1390	0.2886
4. FF MPC	0.0643	0.0178	0.0618

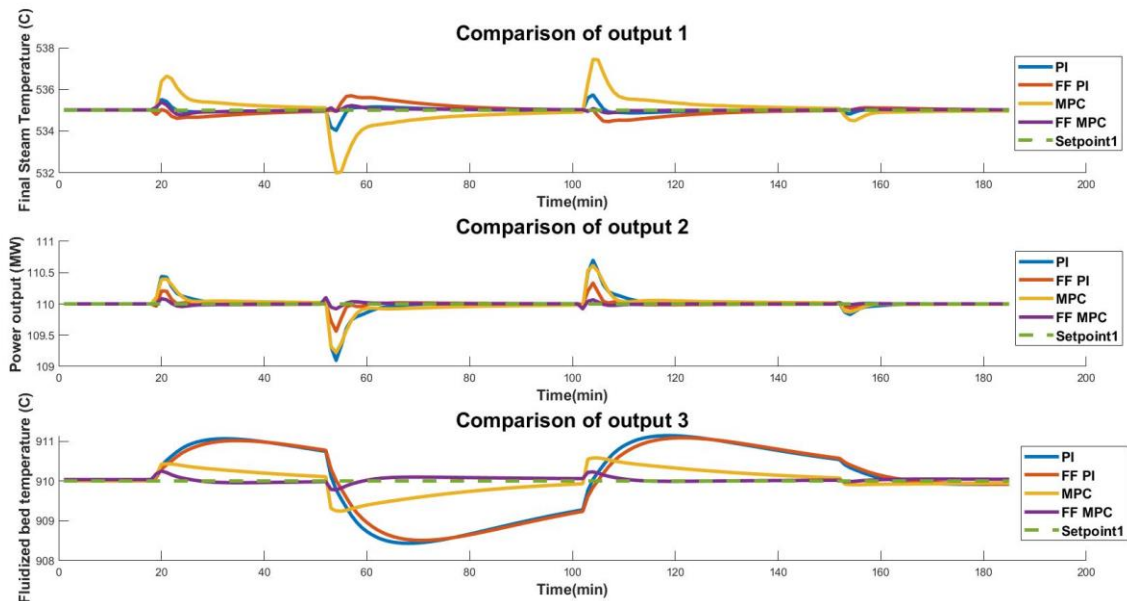**Figure 8.** Control action on Inputs in Scenario 3

Figure 9. Control action on inputs in Scenario 3

5 Conclusions

Four different control strategies were implemented on a dynamic boiler model and compared, based on control and statistical metrics. It was shown that adding feed-forward signal about the fuel moisture, improves the control parameters for both PI and MPC control. By most analyzed parameters FF MPC shows the best performance.

PI FF control provide better results than MPC for the outputs which can be controlled with 1 input manipulation. For the output that depends on most inputs (fluidized bed temperature), MPC show better control results. This shows the limitation of PI control.

FF MPC with accurate characterization of the fuel moisture can help towards dealing with the disturbances in the plant caused by fuel characteristics variation.

Future work is aimed towards exploring the capabilities of MPC with more operation realistic scenarios. Change of power output based on documented operational data, introduce performance deterioration and constraints set based on actuators properties are some of the planned features to be added.

Acknowledgements

References

U. Arena, R. Chirone, M. D'Amore, M. Miccio, and P. Salatino. Some Issues in Modelling Bubbling and Circulating Fluidized-Bed Coal Combustors. *Powder Technology*, 82(3): 301–16, 1995.

K. Atsonios, A. Nesiadis, N. Detsios, K. Koutita, N. Nikolopoulos, and P. Grammelis. Review on Dynamic Process Modeling of Gasification Based Biorefineries and Bio-Based Heat & Power Plants. *Fuel Processing Technology*, 197: 1061-88, 2020. https://doi.org/10.1016/j.fuproc.2019.106188.

K. J. Åström and R. D. Bell. Drum-Boiler Dynamics. *Automatica*, 36: 363-378, 2000.

D. S. Carrasco and G. C. Goodwin. Feedforward Model Predictive Control. *Annual Reviews in* Control, 35(2): 199–206, 2011. http://dx.doi.org/10.1016/j.arcontrol.2011.10.007.

F. Casella, J. G. van Putten, and P. Colonna 2014. DYNAMIC SIMULATION OF A BIOMASS-FIRED STEAM POWER PLANT: A COMPARISON BETWEEN CAUSAL AND A-CAUSAL MODULAR MODELING, *In Proceedings - IMECE2007,2007 ASME International Mechanical Engineering Congress and Exposition November 11-15, 2007, Seattle, Washington, USA,* pages 205-216, 2007.

A. Galgano, P. Salatino, S. Crescitelli, F. Scala, and P. L. Maffettone. A Model of the Dynamics of a Fluidized Bed Combustor Burning Biomass. *Combustion and Flame*, 140(4): 371–84, 2005.

M. Huttunen, J. Peltola, S. Kallio, L. Karvonen, T. Niemi, and V. Y. Outinen. Analysis of the Processes in Fluidized Bed Boiler Furnaces during Load Changes. *Energy Procedia*, 120: 580–87, 2017. http://dx.doi.org/10.1016/j.egypro.2017.07.175.

B. Leckner. Fluidized Bed Combustion Research and Development in Sweden: A Historical Survey. *Thermal Science*, 7(2): 3–16, 2003.

A. Peña. Bubbling Fluidized Bed (BFB) When to Use This Technology?. *In Proceedings - IFSA 2011, Industrial Fluidization, November 2011 South Africa,* pages 1–12, 2011.

J. S. Qin and T. A. Badgwell. A Survey of Industrial Model Predictive Control Technology. *Control Engineering Practice*, 11: 733–64, 2003.

S. Ravelli, A. Perdichizzi, and G. Barigozzi. Description, Applications and Numerical Modelling of Bubbling Fluidized Bed Combustion in Waste-to-Energy Plants. *Progress in Energy and Combustion Science*, 34(2): 224–53, 2008.

R. Saidur, E. A. Abdelaziz, A. Demirbas, M. S. Hossain, and S. Mekhilef. A Review on Biomass as a Fuel for Boilers. *Renewable and Sustainable Energy Reviews*, 15(5): 2262–89, 2011. http://dx.doi.org/10.1016/j.rser.2011.02.015.

J. Sandberg, R. B. Fdhila, E. Dahlquist, and A. Avelin. Dynamic Simulation of Fouling in a Circulating Fluidized Biomass-Fired Boiler. *Applied Energy*, 88(5): 1813–24, 2011. http://dx.doi.org/10.1016/j.apenergy.2010.12.006.

F. Scala and R. Chirone. Fluidized Bed Combustion of Alternative Solid Fuels. *Experimental Thermal and Fluid Science*, 28(7): 691–99, 2004.

P. Szentannai. Modern Power Plant Control for Energy Conservation, Efficiency Increase, and Financial Benefit. In Wei-Yin Chen, Toshio Suzuki and Maximilian Lackner, editors, *Handbook of Climate Change Mitigation and Adaptation*, pages 1–20, 2015. ISBN 978-1-4614-6431-0, doi: 10.1007/978-1-4614-6431-0_22-2

P. Szentannai. Mathematical Modeling and Model-Based Optimum Control of Circulating Fluidized Bed Combustors. *Periodica Polytechnica Civil Engineering*, 55(1): 3–11, 2011.

W. Tan, J. Liu, T. Chen, and H. J. Marquez. Comparison of Some Well-Known PID Tuning Formulas. *Computers and Chemical Engineering*, 30(9): 1416–23, 2006.

L. Wester. *Tables and diagrams for energy technical calculations,* Marklund Solutions, Sweden, 2015.

N. Zimmerman, K. Kyprianidis, and C. F. Lindberg. Waste Fuel Combustion: Dynamic Modeling and Control. *Processes*, 6(11): 1–19, 2018.

Modelling of Water Levels in a Reservoir using Modelica

Willem Mejer   Dietmar Winkler

Department of Electrical Engineering, Information Technology and Cybernetics
University of South-Eastern Norway, Porsgrunn, Norway
`willem.meijer@skagerakenergi.no, dietmar.winkler@usn.no`

Abstract

The regulation of hydro power plants is influenced by amongst others reservoir capacity and the knowledge about how it is influenced by changes in in- and outflow. The hydro power reservoir Aurdalsfjord has a small capacity compared to the inflow, a narrowing between inflow in the north and outflow in the south and has implemented strict governmental restriction.

The paper will present a reservoir model that was developed in order to investigate the water flow and water level conditions in Aurdalsfjord. The water level in each cell is described with a mass balance, while the flow between two cells is described with a momentum balance. The balance equations are formulated as DAEs. The model is developed in Modelica which allows it to be used in connection with models of a wide variety of other components that are present in a hydro power system. The model will be integrated in the open-source hydro power library OpenHPL developed at USN.

The Modelica model shows promising results. For Aurdalsfjord it needed to be adjusted with non-physical parameters. The reservoir model should be implemented for a less complex shaped reservoir in order to see if the challenges are related to the Modelica implementation or Aurdalsfjords shape.

Keywords: modelling, hydro power, reservoir dynamics, Modelica, OpenHPL

1 Introduction

This paper is based on a Master's Thesis (Meijer 2020) from the spring of 2020 at the University of South-Eastern Norway (USN). A dynamic reservoir model was developed with Modelica in order to investigate time-constants between in- and outflow in the hydro power reservoir Aurdalsfjord. The paper will present the Modelica model and use the Aurdalsfjord as an example in order to present some simulation results.

1.1 Hydro Power

Hydro power represents the main energy source in the Norwegian electricity supply and is the most important sustainable energy source worldwide. It transforms the potential and kinetic energy of flowing or falling water into electrical energy.

The regulation of hydro power plants is influenced by electricity prices, inflow, reservoir capacity and governmental restrictions. The relation between reservoir capacity and inflow decides the flexibility of the power plant. If the reservoir is small compared to the inflow, the inflow is more important and the power plant less flexible. If the reservoir however is large compared to the inflow, the amount of stored water is more important.

For some reservoirs these factors prove more challenging than others due to their capacity, shape or strict governmental restrictions. Aurdalsfjord is one example. It is the intake reservoir to the power plant Bagn, which is operated by the Norwegian hydro power company Skagerak Kraft AS. Its capacity is small compared to the inflow (received from two other power plants), has a narrowing between inflow in the north and outflow in the south and has implemented strict governmental restriction.

1.2 Modelica

Modelica is an object-oriented modelling language. It is designed to implement dynamic models equation systems as differential and algebraic equations (DAE). In contrast to most programming languages, Modelica translates the code to its most optimal state in C-code before simulation. Hence, the order in which the equations are written is not relevant. Models from a high variety of engineering fields can be connected, which makes it possible to develop full hydro power plants models.

Modelica models can be used with an optimisation and automation infrastructure using the python application interface OMPython (Open-Source Modelica Consortium 2020a). It is part of the open-source OpenModelica installation (Open-Source Modelica Consortium 2020b).

1.3 Goal and Scope

The dynamic reservoir model is developed that takes the in- and outflows as inputs and gives the water levels as output. The models purpose is to investigate water flows and levels in hydro power reservoirs under different in- and outflow situations. The model dynamics will be described by mass and momentum balances as DAEs.

The model will be integrated into the open-source hydro power library, OpenHPL (TMCC, USN 2020). OpenHPL is developed by USN for teaching and research purposes. The model is developed with OpenModelica, simulated with OMPython and tested with Dymola (Dassault Sys-

tèmes 2020).

2 Mathematical Description

The flow in a reservoir can be described with a combination of mass and momentum balances. The reservoir needs to be discretised into a grid, where the water level in each cell is described with a mass balance and the flow between two cells is described with a momentum balance. Only horizontal flows are taken into account, vertical flows are neglected, as it is assumed that horizontal flow is significantly larger than vertical flow.

The main purpose of the model is to describe the water level in each of the cells. The input parameters are the in- and outflows at the borders. A block diagram for the general model is shown in Figure 1.
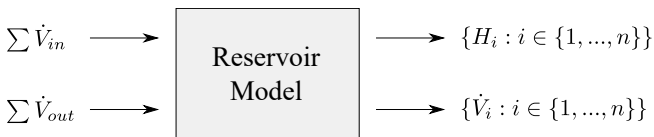

Figure 1. Block diagram for the model.

A DAE is a set of equations containing both ODEs and algebraic equations containing both variables and parameters. The ODEs describe the change of a quantity in time, while the algebraic equations describe the quantities in the ODEs by other variables and/or parameters. In order to be solvable, the DAE needs the same number of equations as variables.

In order to describe the equations, mass balance control volumes and momentum balance control volumes are introduces. The control volumes contain variable (mass or momentum) which changes within a limited volume. The change is caused by forces within the control volume or flow across the borders. The mass balance control volumes are the same as the reservoir cells, while the momentum balance control volumes cover half of one cell and half of the next.

2.1 Mass Balance

The mass balance describes the change of mass in a control volume over time. The change is equal to the mass flow into the control volume subtracted by the mass flow out of the control volume, as shown in the Equation 1. It is assumed that there is no internal mass conversion, as this is only relevant for nuclear reactions.

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \dot{m}_{\text{in}} - \dot{m}_{\text{out}} \tag{1}$$

The goal of the mass balance equation is to find the water level in the control volume as a function of the volume flow. Hence, the mass and mass flows need to be described by the algebraic equations in Equation 2.

$$m = \rho \cdot \Delta x \cdot \Delta y \cdot h$$
$$\dot{m}_{\text{in}} = \rho \dot{V}_{\text{in}} \tag{2}$$
$$\dot{m}_{\text{out}} = \rho \dot{V}_{\text{out}}$$

In the reservoir model, each control volume has borders in both x- and y-direction, as shown in Figure 2. Hence, the mathematical description needs to be modified to fit two flow directions. The *in* and *out* indices are replaced by using positive and negative flow directions. The complete mass balance is described with Equation 3.
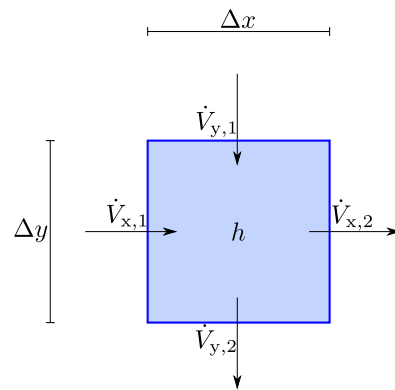
Figure 2. Mass balance control volume seen from above. The arrows show positive flow direction.

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \dot{m}_{\text{x1}} + \dot{m}_{\text{y1}} - \dot{m}_{\text{x2}} - \dot{m}_{\text{y2}}$$
$$m = \rho \cdot \Delta x \cdot \Delta y \cdot h$$
$$\dot{m}_{\text{x1}} = \rho \cdot \dot{V}_{\text{x1}}$$
$$\dot{m}_{\text{x2}} = \rho \cdot \dot{V}_{\text{x2}} \tag{3}$$
$$\dot{m}_{\text{y1}} = \rho \cdot \dot{V}_{\text{y1}}$$
$$\dot{m}_{\text{y2}} = \rho \cdot \dot{V}_{\text{y2}}$$

2.2 Momentum Balance

The momentum balance describes the change in momentum over time in a control volume as a function of momentum flow in and out of the system in addition to the sum of all forces acting on the flow, see Equation 4.

$$\frac{\mathrm{d}M}{\mathrm{d}t} = \dot{M}_{\text{in}} - \dot{M}_{\text{out}} + \sum F \tag{4}$$

In order to connect momentum to volume flow, the algebraic equations in Equation 5 are introduced.

$$M = m \cdot v$$
$$m = \rho \cdot A \cdot \Delta l \tag{5}$$
$$v = \frac{\dot{V}}{A}$$

The momentum flow can be described as a function of volume flow in and out of the system as shown in Equation 6.

$$\dot{M}_{\text{in/out}} = \dot{m}_{\text{in/out}} \cdot v_{\text{in/out}}$$
$$\dot{M}_{\text{in/out}} = \rho \cdot \dot{V}_{\text{in/out}}$$
$$v_{\text{in/out}} = \frac{\dot{V}_{\text{in/out}}}{A_{\text{in/out}}}$$
(6)

The sum of the forces is given by the forces parallel to the horizontal flow. The only two forces which are taken into account are the pressure and the friction forces. It can be shown that they can be given by Equation 7 (pressure force) (Plumb 2014) and Equation 8 (friction force) (Lie 2017).

$$F_{\text{p}} = \rho g \Delta w \frac{h_1^2 - h_2^2}{2}$$
(7)

$$F_{\text{f}} = \Delta l \rho g A S_{\text{fs}}$$
$$S_{\text{fs}} = k_{\text{s}}^2 \dot{V}^2 \frac{\phi}{A^3}$$
(8)

The friction slope, S_{fs} is introduced as part of the friction force. The wetting perimeter, ϕ, is the total circumference of the cross section that produces friction. Due to the model containing mainly cells in the centre of the grid (not without borders), it is assumed that the wetting perimeter is equal to only the width, Δw.

In contrast to the mass balance, the momentum balance is only applicable in one flow direction. Hence, Δw and Δl equals Δx and Δy depending on the direction. Figure 3 shows a control volume for a momentum balance. The complete momentum balance is described with Equation 9. In order to take direction into account, all squared flows are split into a product of the flow times the absolute value of the flow.
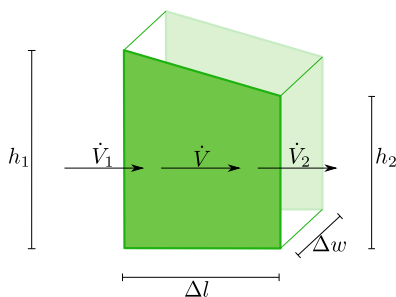

Figure 3. Momentum balance control value.

$$\frac{\mathrm{d}M}{\mathrm{d}t} = \dot{M}_1 - \dot{M}_2 + F_{\text{p}} - F_{\text{f}}$$
$$m = \rho \cdot \Delta l \cdot \dot{V}$$
$$\dot{M}_1 = \rho \frac{\dot{V}_1 |\dot{V}_1|}{A_1}$$
$$\dot{M}_2 = \rho \frac{\dot{V}_2 |\dot{V}_2|}{A_2}$$
(9)
$$F_{\text{p}} = \rho g \Delta w \frac{h_1^2 - h_2^2}{2}$$
$$F_{\text{f}} = \Delta l \rho g \bar{A} S_{\text{fs}}$$
$$S_{\text{fs}} = k_{\text{s}}^2 \dot{V} |\dot{V}| \frac{\phi}{\bar{A}^3}$$

The two sides of the momentum balance control volume can contain different depth values and be redrawn as Figure 4. Due to only horizontal flow being taken into account, the bottom part does *not* influence the system. The heights h_1 and h_2 in the pressure force (Equation 7) need to have the same reference point. Hence, Figure 3 shows only the upper part of the control volume.
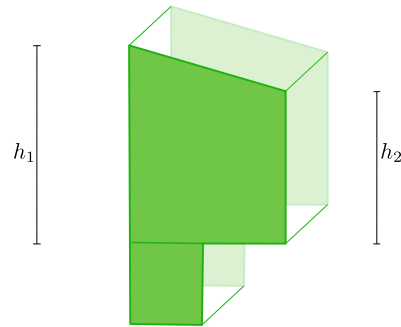
Figure 4. Different depth values.

There can occur situations where the level at one side is lower than the floor on the other side, as shown in Figure 5. In this case there is only flow from the top side. The adjustments in Equation 10 are made in order for Equation 9 to make sense. The reverse can also occur (left side below right side floor).
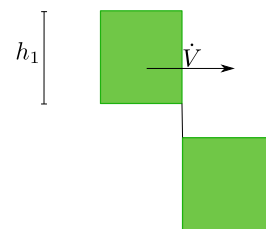
$$\dot{M}_2 = 0$$
$$h_2 = 0$$
(10)

Figure 5. Case where water level on one side is below floor on other side.

2.3 Connecting Mass and Momentum Balance

In order to describe the complete dynamics in the system three *stacked* grids are needed, where the grids contain control volumes for:

- Mass balance

- Momentum balance in x-direction

- Momentum balance in y-direction

The three grids are connected through the mass balance control volumes, which receive mass flows described by both momentum balance grids (x- and y-directions). The momentum balance control volumes receive water levels from the mass balance grid. Figure 6 shows a 1-dimensional version of a stacked grid.
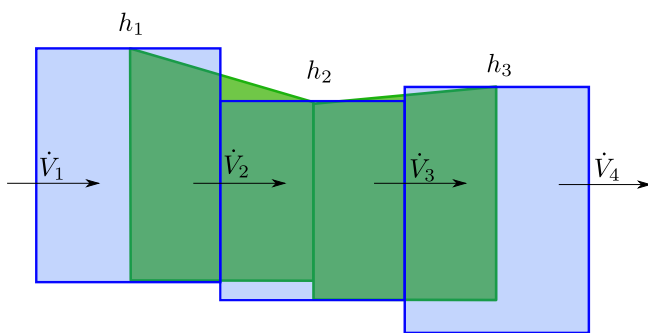
Figure 6. Mass balance (blue) and momentum balance (green) grids stacked on top of each other.

3 Implementing in Modelica

The goal of the reservoir model is to implement it together with in- and outflow components. The implementation of Aurdalsfjord is shown in Figure 7. The in- and outflow models read data from a csv-file. The `data record` (yellow box) contains values for relevant constants (*e.g.*, g, p_a) and is a functionality from OpenHPL.

This section will provide a description of how the main model, `Reservoir`, is implemented. The description is focused on the use of the model and the basic ideas behind the implementation. In addition, it is described how the model can be used to build more complex models, and simulation results are shown.

3.1 Reservoir

The reservoir model describes the water levels in the reservoir by receiving the in- and outflow as the inputs. Due to the model being implemented in Modelica as DAEs, all other values calculated along the way, *e.g.*, volume flows between cells, are also available outputs.

3.1.1 Parameters

The parameters with explanations are shown in Figure 8. Some of the parameters are further explained.

`depth` is an array containing the depth values for the reservoir cells. The amount of cells is also (indirectly)
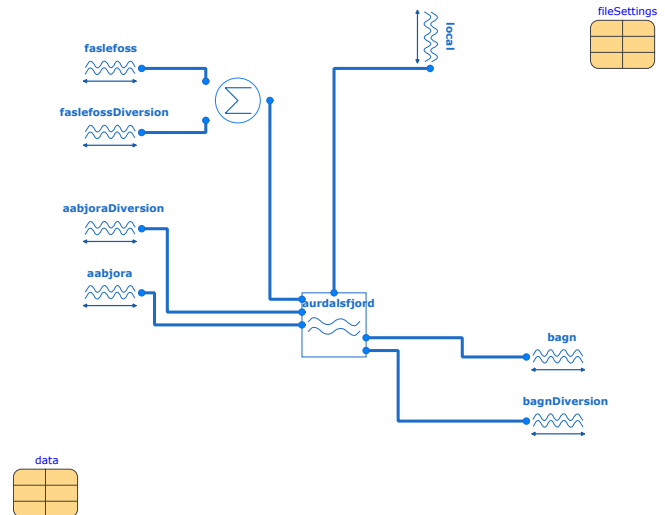
Figure 7. Aurdalsfjord implemented in Modelica.

defined by this parameter. Islands within the grid are possible and need to be given as *zero-cells* (0 m depth). Zero-cells are also possible at the borders, however *only* if the border is not connected to an in- or outflow. Due to the local inflow being distributed over all borders, it is generally recommended to *not* have zero-cells at the borders.

The maximum water level is provided in `H_max`, which must not be misconceived as highest regulated water level (HRWL). The level is given as the maximum water level that can be simulated. It should be set higher than HRWL, so the simulation is able to cover situations with too high water levels.

The parameters `f_p` and `f_M` are factors that influence the weight of the pressure force and the incoming and outgoing momentum flows on the calculations. This is included in order to make the model work better for the Aurdalsfjord case, and is explained further in subsection 3.4. If there is no data to optimise against, the factors should be set to 1.

`H_0` defines the initial water level and applies for the entire reservoir. This was decided due to the model being limited to rectangular shapes. Hence, for more complex reservoirs, several `Reservoir` need to be connected and all of them can be contain `H_0` values.

3.1.2 Basic Principles

The cell structure within the `Reservoir`-model is based on the principles of the stacked grids. Models for the mass balance and for the momentum balance are developed. The models are based on the mathematical formulations in subsection 2.1 and subsection 2.2. In addition, a `boundary`-model is developed to control the in- and outflow of the reservoir.

A `connector` is developed to be able to send four values (water level, volume flow previous and next volume flows) from model to model. The mass balance model receives the volume flow and water levels across all four borders through the `connector` and calculates the wa-
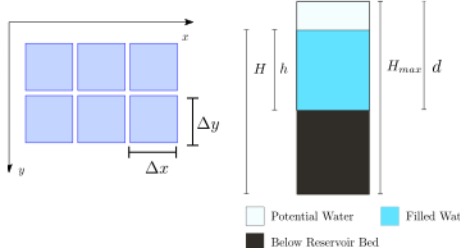
Figure 8. `Reservoir`-parameters.

Figure 9. Grid structure in `Reservoir`.

ter level. The momentum balance model needs the water levels of the connected mass balance models and the momentum flow of the adjacent momentum balance models. The boundary models decide the flow into the rest of the reservoir based on the connected in- and outflows.

The mass balance models are connected at all four borders to momentum balance models, except at the borders of the reservoir. Here, they are connected to boundary-models. Each model in the grid structure calculates either a water level or a volume flow, and is always connected to a model that calculates the opposite. This principle was introduced in order for all models being independent on whether they are connected at a border or in the middle of the reservoir.

Figure 9 shows the inner workings of the `Reservoir` model. The blue circles at the edges are `connectors` which are connected to the inflow, outflow and local inflow (same as in Figure 7). They only send mass flows and pressures, which are *translated* in the boundary-models into water levels and volume flows.

The local inflow is split into equal parts and connected to all boundary-models (dark blue lines). If no in- or outflow is connected, Modelica views it as 0 flow. The inflow and outflow are connected to the parameter defined boundaries.

3.2 Connect Multiple Reservoirs

In cases where the reservoir has non-rectangular shapes, several `Reservoir`-models need to be connected. The models are connected at one of the in- or outflow connectors. Due to each `Reservoir` needing a flow value as the input, it is not possible to connect `Reservoir`-models directly. Another model that calculates the flow between them is necessary.

A `ReservoirConnector` model is developed. It is a simplified version of the momentum balance model. Due to the only values given through the `connector` being mass flow and pressure (can be transformed to volume flow and level), the incoming and outgoing momentum flows are unknown and set to zero. T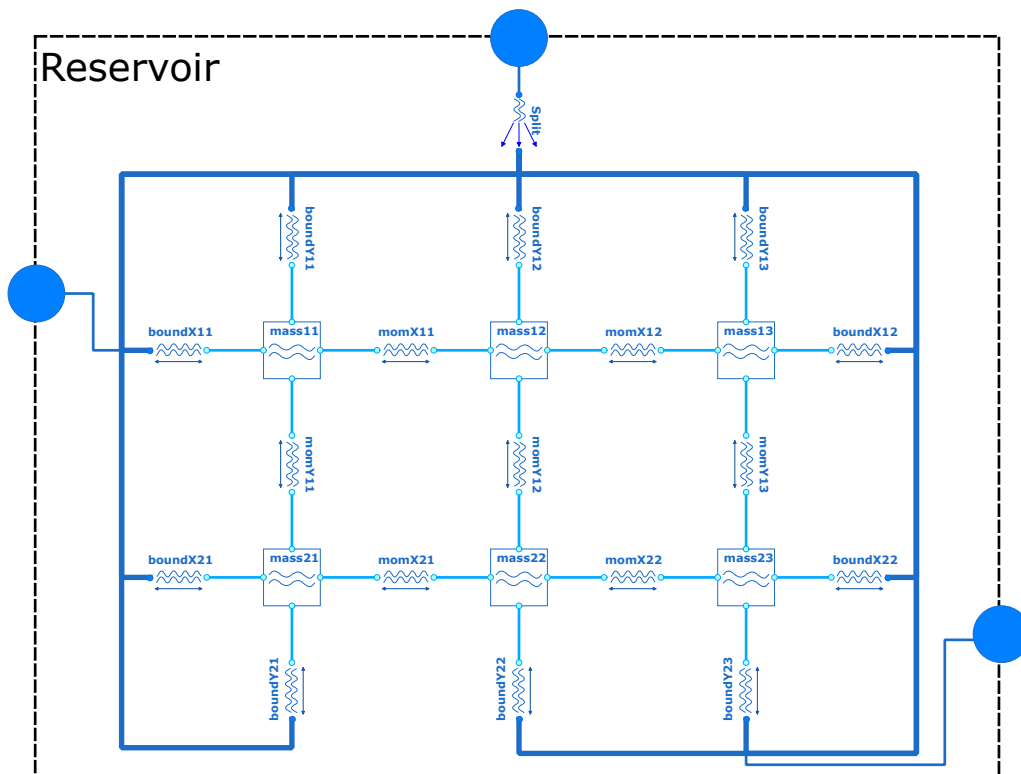he rest of the calculations are identical to the momentum balance. In addition, the two connected cells (from each reservoir) need to have the same depth-value. It is possible to have several parallel connections between the same reservoirs. Otherwise each reservoir connection would introduce a bottleneck in he system. An example of two connected reservoirs is shown in Figure 10.

For Aurdalsfjord a total of 13 `Reservoir`-models are connected. They are shown in Figure 11a. The discretised grid is shown in Figure 11b.

3.3 Hydro Power Plant Model

The developed models can be included into full hydro power plant models together with models from OpenHPL. An example is shown in Figure 12.
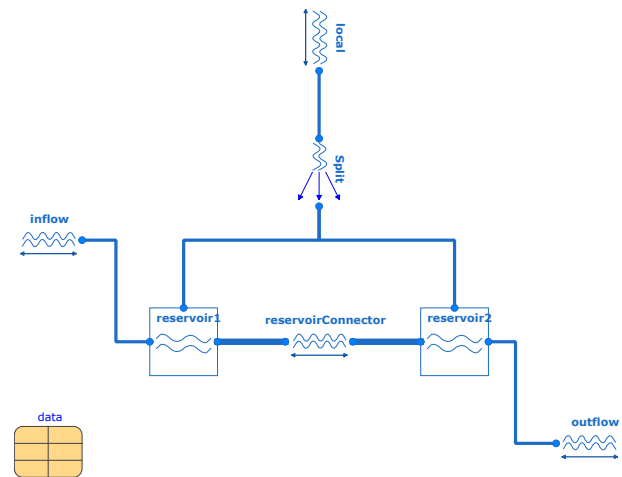
Figure 10. Two connected reservoirs in Modelica.
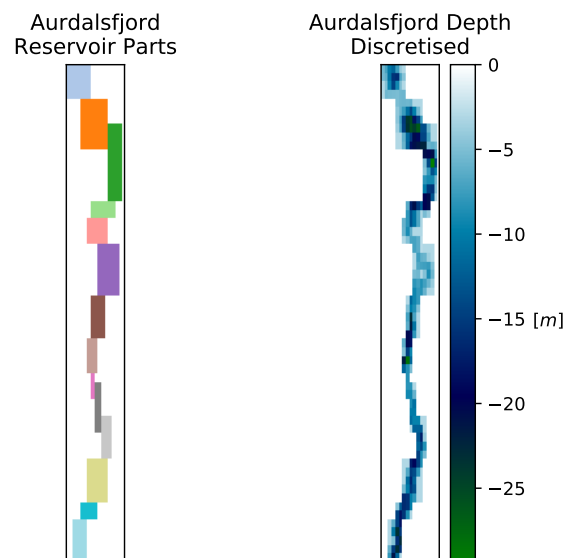
(a) Sub-reservoirs.   **(b)** Cells with depth values.

Figure 11. Aurdalsfjord discretised into sub-reservoirs and cells.

3.4 Simulations with Aurdalsfjord

In the simulation results for Aurdalsfjord, the simulated and measured values for the two measurement stations are compared. One measurement station (*Sundvoll*) is located north of the narrowing, while the other is located at the dam south of the narrowing. All simulations are performed with actual in- and outflow data for winter periods in order to reduce the uncertainty caused by local inflow. The simulation step is set to 3600 s and the tolerance to 0.01. The simulation time for 24 hours is approximately 1 hour.

The initial simulations for Aurdalsfjord proved unsuccessful, due to a spike in water level at the dam location
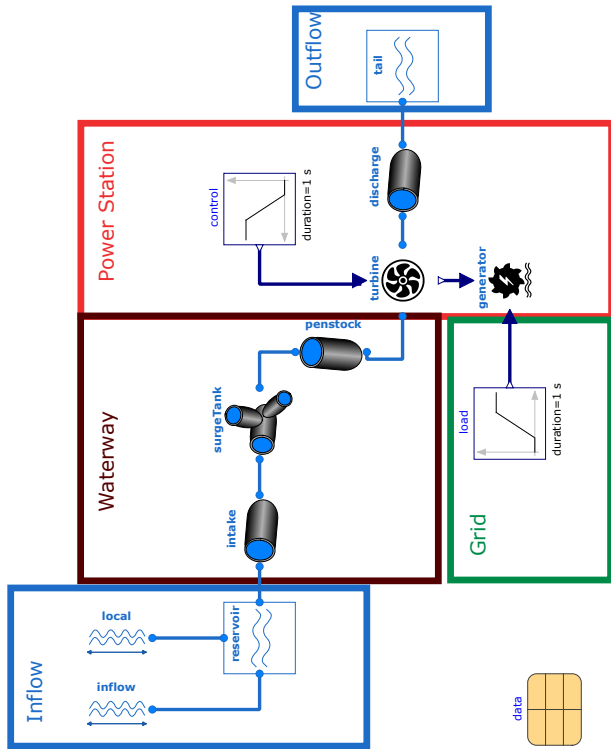
Figure 12. `Reservoir`-model integrated in hydro power plant model with OpenHPL-models.

leading to a simulation breakdown (shown in Figure 13a). This is solved by adding the additional parameters, f_M and f_p to the first equation in Equation 9, resulting in Equation 11.

$$\frac{\mathrm{d}M}{\mathrm{d}t} = f_M \left(\dot{M}_1 - \dot{M}_2 \right) + f_M F_p - F_f \qquad (11)$$

With the additional factors, the parameters k_s, f_p and f_M are estimated to be 0.136, 1.0 and 0.4 respectively. Figure 13b shows the simulation result for the estimation period, while Figure 13c shows the simulation result for a different period using the same parameter values.

4 Discussion

4.1 Implementation of Model

Making a model as detailed and complex as possible is not necessarily better than keeping the model simple. More detailed models include more physical attributes of the actual system, however it will also increase the simulation time. An example is the choice to solely focus on horizontal flow. It would also be possible to include equations that calculate vertical flow in the reservoir, which would be more accurate to reality. The equation system becomes significantly more complex, as things like temperature need to be taken into account. However, water levels are not influenced by vertical flow, and the influence on the horizontal flow is minimal.
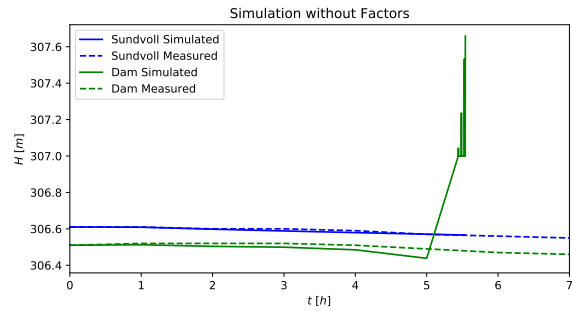
(a) Initial simulation.
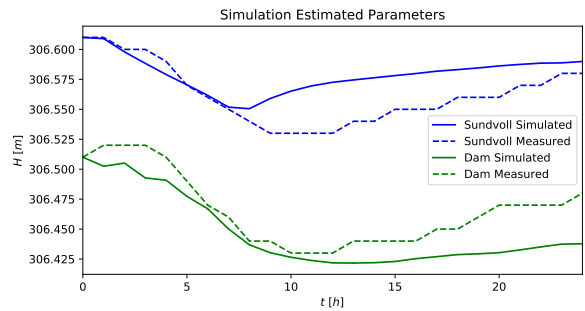
(b) Parameter estimation result.
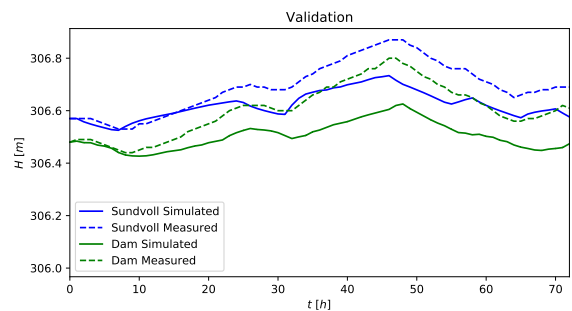
(c) Parameter validation.

Figure 13. Simulations showing simulated and measured values.

In the sum of forces in Equation 6, the pressure and friction forces are included. The gravity and Coriolis forces could also have been included. If the reservoir floor is described by a slope, the gravity that pushes the water down also pushes it partly forward (due to the slope). However, in the developed model, slopes are described as steps. In this case the gravity force only contains a vertical component. The momentum balance control volumes could also have been described with slopes, it was however decided not to include this, as most of the gravity force still only provides vertical flow. The gravity is however still included as part of the pressure force. The Coriolis force always points to the west due to the rotation of the earth. If included, the coordinate system in which the grid is given to the model, would need to be accurate according to earths coordinate system. The force would only be included in the flow in x-direction. The implementation of the force would not overly complicate the system, however due to Aurdalsfjord being narrow it was assumed

that the effect on this exact case was minimal. It could be an option to include this in the future, especially if larger (wider) reservoirs are to be implemented.

Several simplifications are made with the implementation in Modelica, in order to keep the model simple. They are less significant if the chosen grid contains small cells. The two most significant are the connection of mass and momentum grids and the wetting perimeter in the friction force.

The momentum balance grid is on top of the mass balance grid, where each momentum control volume covers half of one mass control volume and half of the next mass control volume. This means that the mass flows in and out of the momentum control volume ($\dot{m}_{in}$ and $\dot{m}_{out}$) should be the same as the mass flowing in the middle of the mass control volumes. In the model, the mass flows in and out of the momentum balance model are set equal to the mass flows in and out of the adjacent mass balance models. Hence, the mass flows are moved by half the length of one mass control volume.

For open channel flows (1-dimensional), the wetting perimeter (ϕ) in the friction force is normally set equal $2h + \Delta w$. The $2h$ part represents the friction against the side walls while the Δw part represents the friction against the reservoir floor. In the model however, the $2h$ part is dropped, due to it being assumed that most cells are in the middle of the reservoir without side walls. This neglects both the cells which are actually at the side of the reservoir and the fact that the reservoir floor is not at the same level through the reservoir. In some of the reservoir the water will flow parallel to a part with less depth and thereby have friction with a (low) side wall.

Instead of basing the model on the DAEs, the model could have been based on the Shallow Water Equations (SWE) (Plumb 2014). The equations are developed from the mass and momentum balance equations, where the cells are made infinitesimal small. Despite the SWE being used in many applications (Behrens 2017) it was chosen to use the DAE. DAEs are normally faster to solve in numeric calculations and Modelica is designed for implementation of DAEs. In addition, it is questionable if using PDEs are significantly more accurate due to larger reservoirs containing many uncertainties.

In OpenHPL there are models available for reservoir and open channel modelling (TMCC, USN 2020). The reservoir was not considered complex enough, due to it not modelling the dynamics within the reservoir (only the flow out). The open-channel models is implemented with the SWE as PDEs for one directional flow. This could have been extended to two directional flow and used as an initial building block, instead of the mass and momentum balances.

4.2 Simulation Results

Figure 13a shows that with both new factors set to 1, the water level at the dam starts spiking the momentum the level increases fast and leads to a breakdown of the simu-

lation. The estimated parameters show that the influence of momentum flow on the momentum balance should be reduced significantly. Both of these statements indicate that the reservoir is responsive to large flows. The fact that only the dam level starts behaving unrealistic also indicates that the narrowing might be part of the problem. A reason might be that the generic model is designed for a reservoir where most cells are in the middle of the reservoir (without borders), while in the narrowing all cells are at the borders.

The estimated parameters show that the pressure factor should stay around 1, while the momentum factor influences the system more significantly. It could be considered to remove the pressure factor from the model, which would force it to 1.

The results in Figure 13b and Figure 13c show that the model is able to simulate realistic results when the parameters are estimated for the same period. However, when the same parameters are used for a different simulation period, the results are less accurate. Due to the time spent on the parameter estimation is limited, it could be possible to improve them and thereby possibly improve the simulation results.

5 Conclusion

The Modelica model shows promising results. For Aurdalsfjord, a reservoir with a narrowing between in- and outflow, it needed to be adjusted with non-physical parameters. They reduce the weight of the pressure force and the momentum flows in the momentum balance. The reservoir model should be implemented for a less complex shaped reservoir in order to see if the challenges are related to the Modelica implementation or Aurdalsfjords shape.

The new reservoir model presented in this paper will be part of a future version of OpenHPL.

References

Jörn Behrens. 2017. *Surprisingly Rich in Applications: The Shallow Water Equations and Numerical Solution Methods.* URL: `https://sinews.siam.org/Details-Page/surprisingly-rich-in-applications-the-shallow-water-equations-and-numerical-solution-methods` [Accessed 16 January 2020].

Dassault Systèmes. *Dymola.* 2020. URL: `https://www.3ds.com/products-services/catia/products/dymola/` [Accessed 9 June 2020].

Bernt Lie. *Modelling of Dynamic Systems*, 247 pages, 2017.

Willem Meijer. *Dynamic Reservoir Model Using Modelica: Modelling of Water Levels in Aurdalsfjord.* M.Sc.Thesis, University of South-Eastern Norway, Porsgrunn. 121 pages, 2020.

Open-Source Modelica Consortium. *OMPython - A Python Interface to OpenModelica Communicating via CORBA or ZeroMQ.* 2020a. OpenModelica.

URL: `https://github.com/OpenModelica/OMPython` [Accessed 24 February 2020].

Open-Source Modelica Consortium. *OpenModelica.* (2020b) URL: `https://openmodelica.org` [Accessed 24 February 2020].

Alan Plumb. (2014). *Dynamics of the Atmosphere - Lecture Notes. Chapter 2: The Shallow Water Equations.* Massachusetts Institute of Technology. URL: `http://www-aps.mit.edu/~rap/courses/12333_notes/A2%20SWeqs.pdf` [Accessed 15 February 2020].

TMCC, USN. *OpenHPL - an Open-Source Hydropower Library That Consists of Hydropower Unit Models.* 2020 URL: `https:openhpl.simulati.no` [Accessed 24 February 2020].

The influence of surge tanks on the water hammer effect at different hydro power discharge rates

Madhusudhan Pandey, Bernt Lie

TMCC, University of South-Eastern Norway, Bernt.Lie@usn.no

Abstract

This paper provides an overview of different types of surge tanks used in hydropower systems. The water mass oscillation inside the simple, sharp orifice type, throttle valve, and air-cushion surge tanks are studied. It is found that the diameter of the sharp orifice and the throat plays an important role in obstructing water mass flowing inside the surge tank which consequences to reduce the effect of water hammer over times in the pressure tunnels. Sharp orifice type surge tanks are more efficient to reduce the allowed maximum height of surge tank for avoiding water spilling out of the surge tank during the total load rejection from the prime movers. However, throttle valve surge tanks are more efficient for decaying of pressure surges sooner. It is also found that the difference-amplitude of water mass oscillation inside the air-cushion surge tank is insignificant. Conclusions are drawn based on the case study of Trollheim and Torpa hydroelectric plants in Norway.

Keywords: water mass oscillation, surge tanks throttling, sharp orifice type surge tank, air-cushion surge tank, throttle valve surge tank, water hammer

1 Introduction

1.1 Background

A high-head reaction-turbine hydro power system basically consists of an intake tunnel via a high-pressure steep penstock tunnel to the reaction turbines (eg., Francis turbine). A surge tank is usually placed between the intake pressure tunnel and the penstock. In case of a load rejection[1], the turbine valve is rapidly positioned for a required volumetric flow (discharge) of water through the turbine. During rapid closing of the turbine valve, the water masses flowing in the intake tunnel and in the penstock are suddenly decelerated. A high-pressure region is created at the lower end of the penstock because of the obstructed water-inertia[2] which causes pressure waves to travel in the up-

ward direction[3]. The magnitude of the travelled pressure wave after sudden closure of the turbine valve is termed as a *water hammer*. The energy of the pressure wave is released at the nearest low-pressure free water surface, i.e, at the surge tank placed between the intake tunnel and penstock (Mosonyi, 1991, p. 129).

In this regard, it is of interest to see the effect of the water hammer at different discharges through the turbine during the load acceptance or rejection. The water inside the surge tank oscillates after the energy from the pressure wave is released at the free water surface inside the surge tank. The oscillation of water mass lasts until the pressure wave energy is fully dissipated. The design height and length of the surge tank should thus depend on the amplitude of the pressure wave, i.e., the water hammer. The amplitude of water mass oscillation inside the surge tank can be decreased using water flow-obstruction in the inlet of the surge tank, eg., in case of throttle valve surge tank and sharp orifice type surge tank (Aronovich et al., 1970). Similarly, energy from the pressure wave can be dissipated using pressurized air inside a closed surge tank, usually referred to as an air-cushion surge tank (Vereide et al., 2014). This paper will mainly focus on the simulated response at different discharge for *manifold pressure*[4], velocity, mass flow rate and water mass oscillation inside the different kinds of surge tanks.

1.2 Previous studies

A detailed overview of the time evolution of water mass oscillation inside a surge tank is given in (Guo et al., 2017) with differential equations governing the oscillation phenomenon. Similarly, a law governing oscillation phenomenon inside the simple surge tank is explored in (Travaš, 2014). The water mass oscillation control analysis using a self-adaptive auxiliary control system in the surge tank has been done in (Wan et al., 2019). The solution of water mass oscillation mathematical equations has been done using the finite element method in (Wan et al., 2019).

---

[1] Load rejection is simply a phenomenon where load connected to a prime mover, for eg., Francis turbine, is suddenly disconnected or decreased. However, in case of a load acceptance, a load is connected to the prime mover. A load is anything which is operated with the help of prime mover. For a hydroelectric plant, loads are electrical units connected through the grid in an interconnected electrical network.

[2] The obstructed water mass flowing through the pressure tunnel is generally called as water inertia.

[3] The pressure wave traveled from higher pressure to lower pressure region and dissipated near to free water surface.

[4] It is a bottom pressure point of the surge tank where the outlet of the intake tunnel and inlet of a steep penstock meet.
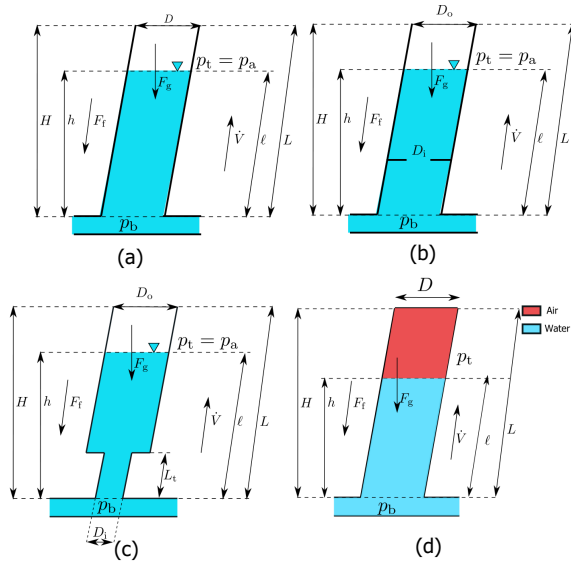
Figure 1. Different types of surge tanks. (a) Simple surge tank without hydraulic resistance. (b) Sharp orifice type surge tank with hydraulic resistance of horizontal bars forming an orifice of a diameter D_i. (c) Throttle valve surge tank with hydraulic resistance of diameter D_i at the entry of surge tank with square expansion from diameter D_i to diameter D_o. The length of the throat is L_t. (d) Air-cushion surge tank filled with air at pressure p_t and diameter D.

1.3 Outline of the paper

The paper is organized by providing a brief introduction to different types of surge tanks and their operation in Section 2. Section 3 provides the simulated responses for Trollheim and Torpa hydro power plants with different types of surge tanks at different discharges. Results and discussions are provided in Section 4 while conclusions and future works are explained in Section 5.

2 Surge tanks and their operation

A detailed mechanistic model of simple, sharp orifice type, throttle valve, and air-cushion surge tank are articulated in (Pandey and Lie, 2020, Submitted) for a Modelica[5] based hydro power library- OpenHPL[6]. OpenHPL is an open-source hydropower library consisting of models for hydropower components that are developed based on mass and 1D momentum balance. It consists of mechanistic models for the flow of water in filled pipes (inelastic and elastic walls, incompressible and compressible water), a mechanistic model of a Francis turbine (including design of turbine parameters), friction models, etc.

The different types of surge tanks are shown in Figure 1. For a simple surge tank shown in Figure 1 (a), during the load acceptance/rejection, a high-pressure region is created at the end of the penstock and at the end of the turbine. The high pressure region thus creates pressure wave which traveled through the penstock releasing pres-

sure wave energy by the means of water mass oscillation inside the surge tank. The height and length of surge tank thus depends on the water mass oscillation inside the surge tank. For a simple surge tank, the maximum height of surge tank would be sum of piezometric height from surge tank bottom to resorvoir surface and the highest amplitude of water mass oscillation during a *total load rejection*[7].

If the height of surge tank is not practically possible then other surge tanks with hydraulic resistances like horizontal bars forming a sharp orifice as in sharp orifice type surge tank or a throat in the entry of surge tank as in throttle valve surge tank can be used. Figure 1 (b) shows a sharp orifice type surge tank with orifice diameter D_i which obstructs water mass moving from the base of surge tank towards the free water surface inside the surge tank. This will cause the oscillation of water mass to dies out sooner than in the simple case. Similarly, the highest amplitude of water mass oscillation is decreased which decreases the practical height of the surge tank. The throat with diameter D_t and length L_t, in case of the throttle valve surge tank as shown in Figure 1 (c), has the same operation as that of sharp orifice type surge tank. Figure 1 (d) shows air-cushion surge tank.

3 Simulated Responses

3.1 Case study: Trollheim HPP

The case study for the simulated responses for different types of surge tanks at different discharge rates is studied for Trollheim and Torpa hydro power plant. The general layout diagram is shown in Figure 2.

3.1.1 Total Load Rejection (TLR)

First, we consider a case of a simple surge tank for Trollheim Hydro Power Plant (HPP) for a layout shown in Figure 2 (a) Trollheim HPP for a total load rejection. Assuming frictionless intake pressure tunnel and ideal gate valve for turbine (i.e., time of opening and closure of the gate valve is *zero*), the maximum allowable height of a simple surge tank for restriction of water-spilling from surge tank is given by the expression as in Eq. 1,

$$H_{ST} = H_{res} + H_{in} + Y_{max}, \tag{1}$$

where Y_{max} is the maximum surge or maximum water mass oscillation height during total load rejection (Mosonyi, 1991, p. 162) given as in 2,
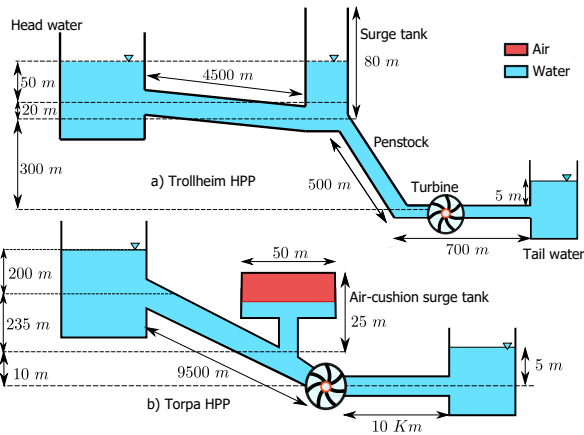
$$Y_{max} = \frac{\dot{V}_n}{A_{in}}\sqrt{\frac{L_{in}}{g}\left(\frac{A_{in}}{A_{ST}}\right)}, \tag{2}$$

where H_{ST}, H_{in}, and H_{res} are height difference for surge tank, intake and resorvoir, respectively. A_{in} and L_{in} are

Figure 2. Layout diagram for Trollheim (Vytvytskyi and Lie, 2019) and Torpa Hydro Power Plant (HPP) (Vereide et al., 2014). Nominal head, nominal discharge, and nominal power output are 370 m, 40 m^3/s and 130 MW for Trollheim HPP, and 445 m, 35 m^3/s and 150 MW for Torpa HPP. Torpa HPP has two turbine units each having nominal power output of 75 MW. The air-cushion surge tank for Torpa HPP has air volume of 13,000 m^3 initially pressurized at 4.1 Mpa. For Trollheim HPP, the diameter for both of the penstock and the surge tank is 4 m while for both of the headrace and the tailrace tunnel is 6 m. Similarly, for Torpa HPP, the diameter of both of the headrace and the tailrace tunnel is 7 m.

cross-sectional area and length of intake pressure tunnel, respectively. $\dot{V}_n$ is the nominal discharge with g as the acceleration due to gravity. From Figure 2 (a) Trollheim HPP we have $H_{res} = 50$ m, $H_{in} = 20$ m and Y_{max} is calculated using expression Eq. 2 as 45 m. Thus, the height of surge tank for avoiding water spilling out for a simple surge tank for Trollheim HPP during total load rejection is 115 m.

Figure 3 shows the turbine valve signal creating a total load rejection at 1500 s and plots of water mass oscillation for simple, sharp orifice type and throttle valve surge tank. It shows that hydraulic resistances in case of sharp orifice type and throttle valve surge tank dampens out the mass oscillation sooner than that of the simple surge tank and the maximum allowed height of surge tank H_{ST} for avoiding water spilling out of surge tank is less for sharp orifice type surge tank during TLR.

3.1.2 Effect of diameter of orifice and throat for TLR

The maximum allowed height of sharp orifice type and throttle valve surge tank for avoiding water spilling through the surge tank can be decreased based on decreasing diameter of orifice and throat as shown in Figure 4.

3.1.3 Total Load Acceptance (TLA)

A case of a total load acceptance is created using turbine guide valve control signal $u_v = \begin{cases} 0.01 & 0 < t \leq 200 \text{ s} \\ 1 & t > 200 \text{ s} \end{cases}$ at time 200 s for Trollheim HPP. The simulated response for water mass oscillation for simple, sharp orifice and throttle
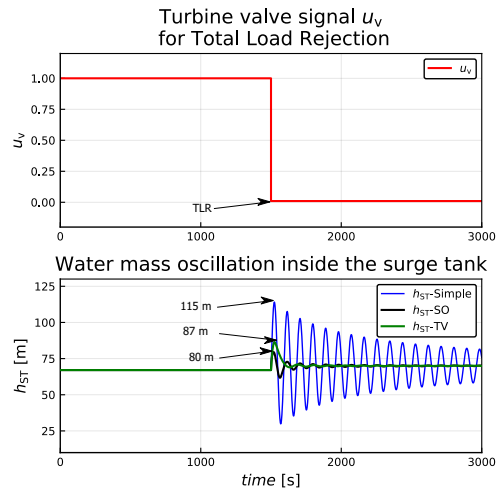
Figure 3. Water mass oscillation inside the surge tank for Trollheim HPP. A total load rejection is created using control signal
$$u_v = \begin{cases} 1 & 0 < t \leq 1500 \text{ s} \\ 0.01 & t > 1500 \text{ s} \end{cases}$$
at time 1500 s. In the figure, TLR repesents total load rejection, SO and TV depicts sharp orifice type and throttle valve surge tank. The maximum amplitude of water mass oscillation h_{ST} is 115 m at around 1500 s for simple surge tank. While for sharp orifice type and throttle valve surge tank it is 80 m and 87 m, respectively. The diameter of orifice for sharp orifice type surge tank D_{so} and that of throat for throttle valve surge tank D_t are both 1 m. The length of throat for throttle valve surge tank is 20 m.
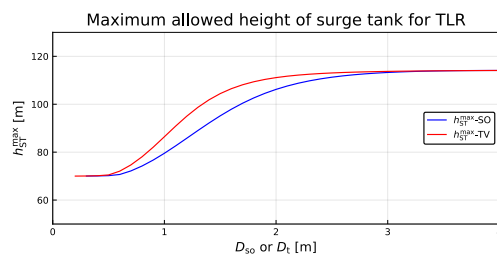
Figure 4. Maximum allowed height of surge tank for different diameter of sharp orifice (SO) type and throttle valve (TV) surge tank. h_{ST}^{max} represent the maximum amplitude of water mass oscillation during TLR. As the diameter of hydraulic resistances like sharp orifice or throat at the entry of the surge tank is decreased the maximum height of water mass oscillation decreased. For example when D_t and D_{so} both are 1 m, h_{ST}^{max} for sharp orifice type surge tank is 80 m and for throttle valve surge tank is 87 m.
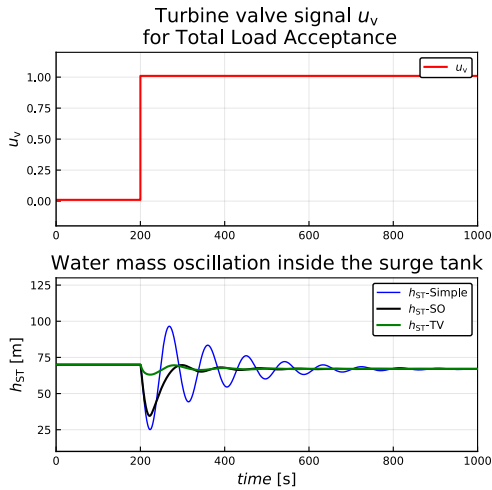
Figure 5. Water mass oscillation inside the surge tanks for TLA. The oscillation dies out soon in case of both sharp orifice and throttle valve surge tank.

valve surge tank is shown in Figure 5.

3.1.4 Partial Load Rejection (PLR)

Partial load rejections and acceptances can be created by changing the turbine's gate signal.

For a 25 % load rejections while the hydropower plant is running at total load the turbine gate signal is generated as,

$$u_{\mathrm{v}} = \begin{cases} 1 & 0 < t \leq 200 \text{ s} \\ 0.75 & t > 200 \text{ s} \end{cases},$$

where the plant is running at total load up to 200 s and with partial load (75 %) after 200 s.

Similarly, for a 50 % load rejection the turbine's gate signal is generated as,

$$u_{\mathrm{v}} = \begin{cases} 1 & 0 < t \leq 200 \text{ s} \\ 0.50 & t > 200 \text{ s} \end{cases},$$

and for for a 75 % load rejection,

$$u_{\mathrm{v}} = \begin{cases} 1 & 0 < t \leq 200 \text{ s} \\ 0.25 & t > 200 \text{ s} \end{cases}.$$

Figure 6 shows water mass oscillation inside the simple, sharp orifice and throttle valve surge tank during the partial load rejections.

3.1.5 Partial Load Acceptance (PLA)

For a 25 % load acceptance while the hydropower plant is running at no load condition, the turbine gate signal is generated as,

$$u_{\mathrm{v}} = \begin{cases} 0 & 0 < t \leq 200 \text{ s} \\ 0.25 & t > 200 \text{ s} \end{cases},$$
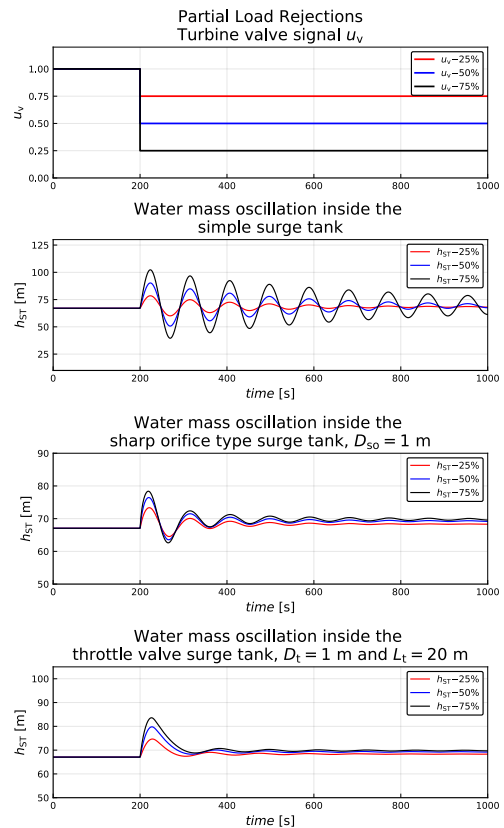
Figure 6. Water mass oscillation inside the surge tanks for PLR. In the figure, $u_{\mathrm{v}} - 25\%$ represents the gate signal for a partial load rejection of 25% of the total load capacity of the plant. Similarly, $h_{\mathrm{ST}} - 25\%$ represents water mass oscillation for a load rejection of 25%.

where as for a 50 % load acceptance,

$$u_{\mathrm{v}} = \begin{cases} 0 & 0 < t \leq 200 \text{ s} \\ 0.50 & t > 200 \text{ s} \end{cases},$$

and for a 75 % load acceptance,

$$u_{\mathrm{v}} = \begin{cases} 0 & 0 < t \leq 200 \text{ s} \\ 0.75 & t > 200 \text{ s} \end{cases}.$$

Figure 7 shows water mass oscillation inside the simple, sharp orifice and throttle valve surge tank during the partial load rejections.

3.2 Case study: Torpa HPP

The water mass oscillation and the air pressure inside the air-cushion surge tank during load rejections and acceptance for Torpa HPP is shown in Figure 8 and 9, respectively.

4 Results, and Discussions

For Trollheim HPP, from Figure 3 in case of a TLR, the maximum allowed height of the surge tank for restriction of water spilling out of a simple surge tank is 115 m. Similarly, for sharp orifice type surge tank it is 80 m and for
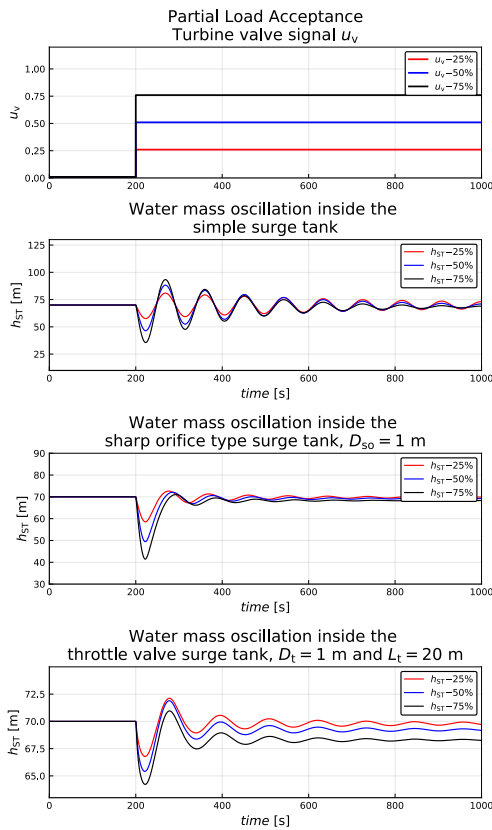
Figure 7. Water mass oscillation inside the surge tanks for PLA. In the figure, $u_v - 25\%$ represents the gate signal for a partial load acceptance of 25% of the total load capacity of the plant. Similarly, $h_{ST} - 25\%$ represents water mass oscillation for a load acceptance of 25% from a no load condition.

Figure 8. Water mass oscillation and air pressure inside the air cushion surge tank for load rejections. In figure, TLR represents a total load rejection.

throttle valve surge tank it is 87 m. From Figure 4 it can be seen that the maximum allowed height of surge tank in case of a total load rejection is decreased as the diameter of sharp orifice and diameter of the throat is decreased. For a surge tank of diameter 4 m, in case of Trollheim HPP, the maximum allowed height of the simple surge tank, h_{ST}^{max} during TLR is same for sharp orifice type surge tank with $D_{so} \in [3,4]$, however, h_{ST}^{max} decreases as $D_{so} \in [0.5,3)$. Similarly, in case of throttle valve surge tank h_{ST}^{max} is same for $D_t \in [2,4]$ and simple surge tank, however, h_{ST}^{max} decreases as $D_t \in [0.1,2)$.

For Torpa HPP, from Figure 8 and 9 in case of load rejections and acceptance, respectively, manifold pressure inside the surge tank does not vary much in case of load acceptance than in case of rejections.

5 Conclusions

The maximum allowed height of a simple surge tank, considering the TLR operation of the plant, can be decreased using a suitable diameter of the sharp orifice in case of a sharp orifice type surge tank and with a throttle valve surge tank with suitable diameter of the throat. The maximum allowed height of the surge tank is lowest in case of sharp orifice type surge tank, however, the mass oscillation dies

out soon in case of throttle valve surge tank with an inference that impact of water hammer in the pressure tunnel is less in case of throttle valve surge tank.

In the case of a sharp orifice type surge tank, the maximum allowed height of the surge tank in comparison with a simple surge tank decreases exponentially for $D_{so} \leq 0.5 \cdot D$ where D and D_{so} are the diameter of the simple surge tank and the diameter of the sharp orifice. Similarly, for throttle valve surge tank $D_t \leq 0.375 \cdot D$ where D_t is the diameter of the throat. Both for load rejections and acceptance, mass oscillation inside the surge tank dies sooner in case of a throttle valve surge tank. The frequency of water mass oscillation in the case of a simple surge tank is the same for both load rejections and acceptance.

For the air-cushion surge tank, water mass oscillation inside the surge tank is insignificant for both load acceptance and rejections. The varying of air pressure inside the surge tank for partial load rejections is greater than that for the partial load acceptance.

References

Grigorii Vladimirovich Aronovich, Nikolai Archilovich Kartvelishvili, and Ya K Lyubimtsev. Water hammer and surge tanks. Technical report, Israel program for scientific translations, 1970.

Junke Guo, Kokob Woldeyesus, Jianmin Zhang, and Xiaoming Ju. Time evolution of water surface oscillations in surge tanks. *Journal of Hydraulic Research*, 55(5):657–667, 2017.
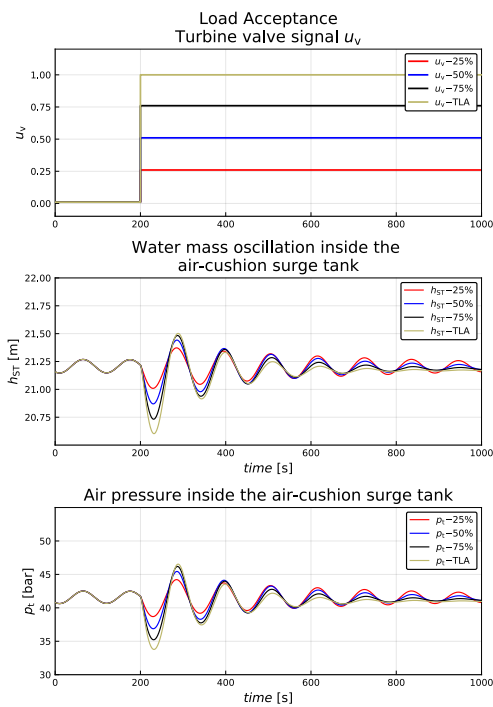
Figure 9. Water mass oscillation and air pressure inside the air cushion surge tank for load acceptance. In figure, TLA represents a total load rejection.

Emil Mosonyi. *Water power development*, volume 2A. Publishing house of the Hungarian Academy of Sciences, 1991.

Madhusudhan Pandey and Bernt Lie. Mechanistic modeling of different types of surge tanks and draft tubes for hydropower plants. In *Proceedings of The 61th SIMS Conference on Simulation and Modelling SIMS 2020, September 22-24, Oulu, Finland*. Linköping University Electronic Press, 2020, Submitted.

Vanja Travaš. Water mass oscillations in a generic surge chamber. *Građevinar*, 66(04.):323–334, 2014.

K Vereide, L Lia, T Nielsen, et al. Physical modelling of hydropower waterway with air cushion surge chamber. In *11th National Conference on Hydraulics in Civil Engineering & 5th International Symposium on Hydraulic Structures: Hydraulic Structures and Society-Engineering Challenges and Extremes*, page 134. Engineers Australia, 2014.

Liubomyr Vytvytskyi and Bernt Lie. OpenHPL for Modelling the Trollheim Hydropower Plant. *Energies*, 12(12):2303, 2019.

Wuyi Wan, Boran Zhang, Xiaoyi Chen, and Jijian Lian. Water hammer control analysis of an intelligent surge tank with spring self-adaptive auxiliary control system. *Energies*, 12 (13):2527, 2019.

Mechanistic modeling of different types of surge tanks and draft tubes for hydropower plants

Madhusudhan Pandey, Bernt Lie

TMCC, University of South-Eastern Norway, Bernt.Lie@usn.no

Abstract

OpenHPL is an open-source hydropower library consisting of models for hydropower components that are developed based on mass and 1D momentum balance. It consists of mechanistic models for the flow of water in filled pipes (inelastic and elastic walls, incompressible and compressible water), a mechanistic model of a Francis turbine (including design of turbine parameters), friction models, etc. This paper includes an extension of OpenHPL with mechanistic models of different types of surge tanks (sharp orifice type, throttle valve surge tank, air-cushion surge tank) and draft tubes (conical diffusers and Moody spreading pipes). The simulated response of the models is presented using a case study of real hydro power plants.

Keywords: surge tanks, draft tubes, air-cushion surge tank, throttle valve surge tank, conical diffuser, Moody spreading pipes

1 Introduction

1.1 Background

The electricity generation from renewables has increased because of the rise in coal prices, oil insecurity, climatic concern (Brown, 2012), and the nuclear power debate (Wikipedia, 2019). There is a demand for renewable-sources economy over the coal-fired economy (Brown, 2012). The renewable energy sources are a combination of intermittent and dispatchable energy sources. Intermittent sources like solar, wind, and tidal power plants exhibit fluctuating power production that creates an imbalance between generation and load. In this regard, renewable dispatchable sources like hydro power plants play a significant role in balancing out the variability caused by intermittent sources. Current hydropower modeling, design, and analysis tools are limited and available commercially. Freely available tools include CASiMiR-Hydropower[1], LVTrans[2], and OpenHPL[3], while commercial tools include Alab[4] and Modelon Hydro Power Library (HPL)[5]. In this regards, it drives motivation for open-source hydro power library development for modeling, design, and analysis.

1.2 Previous studies

A mechanistic model of hydropower systems has been developed in (Splavska et al., 2017) using mass and 1D linear momentum balance which leads to a Modelica[6] based open-source hydropower library OpenHPL, and was initiated in a PhD study (Vytvytskyi, 2019). OpenHPL is under development at the University of South-Eastern Norway. Currently, OpenHPL has units for flow of water in filled pipes (inelastic and elastic walls, incompressible and compressible water) (Vytvytsky and Lie, 2017), a mechanistic model of a Francis turbine (including design of turbine parameters), friction models, etc (Vytvytskyi and Lie, 2018). The library also has draft models for a Pelton turbine, Francis turbine friction model, surge shaft, open channel flow, and a hydrology model. In addition, some accompanying work on analysis tools has been developed in scripting languages (Python, Julia) related to state estimation, structural analysis, etc (Vytvytskyi and Lie, 2019b). The library has been tested on real power plant data (Vytvytskyi and Lie, 2019a). The library is designed to interface to other Modelica libraries, e.g., libraries with generator models, electric grid, etc., for example, OpenHPL can be integrated with PVSystems[7] for photovoltaics as in (Pandey and Lie, 2020).

In this regard it is of interest to further develop units for OpenHPL. This paper primarily focuses on mechanistic models of surge tanks and draft tubes. The simple surge tank mechanistic model developed in (Splavska et al., 2017) is further enhanced by a sharp orifice type surge tank and a throttle valve surge tank considering hydraulic resistance in the inlet to the surge tank. The surge tank model is also further enhanced using air-cushion surge tank as a closed surge tank mechanistic model. A further extension to the library includes mechanistic models of draft tubes: conical diffuser and moody spreading pipes.

1.3 Outline of the paper

The paper is organized as follows. Model developments for surge tanks and draft tubes are provided in Section 2. The simulated response for the developed mechanistic model for surge tanks and draft tubes are presented in Section 3. Conclusions and future work are sketched in

---

[1]http://www.casimir-software.de/save_download.php?language=2

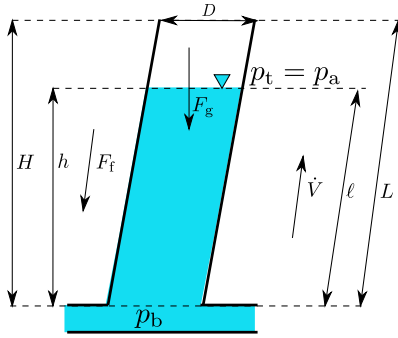[2]http://svingentech.no/about%20lvtrans.html

[3]https://github.com/simulatino/OpenHPL

[4]http://www.alab.no/Alab-Hydropower-Software-Functionality-Alab-Hydropower-Software/Operation-simulation-with-waterway

[5]https://www.modelon.com/library/hydro-power-library/

[6]https://www.modelica.org/

[7]https://github.com/raulrpearson/PVSystems

Figure 1. Simple surge tank with geometrical dimensions of height H, length L, and diameter D. The height of liquid level inside the surge tank is h with slanted length to be ℓ. The volumetric flow rate inside the surge tank is $\dot{V}$ with F_f as fluid friction. F_g is the gravitational force due to fluid mass m inside the surge tank. p_b is the bottom pressure and p_t is the top pressure equivalent to atmospheric pressure p_a for a free-fluid surface. For a hydropower system, we considered fluid inside the surge tank to be water with color as blue.

Section 4.

2 Model Developement

Two main assumptions were made while developing models for hydro power units. First, we consider incompressible water flow inside the units. Second, we consider the inelastic pipe for modeling all types of surge tanks and draft tubes.

2.1 Surge tanks

A surge tank is usually placed between an intake and a penstock pipe in a hydro power system. The prime benefit of a surge tank is to provide a low-pressure region to dissipates pressure energy released during the sudden opening and closing of the turbine valve. Depending on the location and head, the surge tank can be of open type (water surface at atmospheric pressure) or closed type (water surface in contact with pressurized gas/air).

2.1.1 Simple surge tank

A simple surge tank is shown in Fig. 1.

The mass and momentum balance for a dynamical system can be expressed as in (Lie, 2017a, p. 87-88, 226-227) ,

$$\frac{dm}{dt} = \dot{m}_i - \dot{m}_e$$
$$\frac{d\mathcal{M}}{dt} = \dot{\mathcal{M}}_i - \dot{\mathcal{M}}_e + F,$$

where subscript i and e refers to influent and effluent properties, respectively. m, $\mathcal{M}$ and F represents mass, linear momentum and force acted-upon in a dynamical system. If $\dot{m}$ and $\dot{\mathcal{M}}$ are mass flow rate and momentum flow rate for a system with single entry and single exit, it is commonly written as $\dot{m}_i - \dot{m}_e = \dot{m}$ and $\dot{M}_i - \dot{\mathcal{M}}_e = \dot{\mathcal{M}}$.
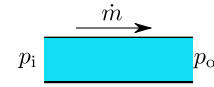
Figure 2. A hydropower unit, for example a simple rigid pipe, with ideal fluid (water) flow from a inlet pressure point p_i to outlet pressure p_o with $\dot{m}$ as *a flow* variable. The pressure point inside the pipe is considered to be *a across* variable.

For surge tanks, if $\dot{m}$ and $\dot{\mathcal{M}}$ represents water mass flow rate and momentum flow rate of water inside the surge tank, respectively, then,

$$\frac{dm}{dt} = \dot{m} \tag{1}$$

$$\frac{d\mathcal{M}}{dt} = \dot{\mathcal{M}} + F. \tag{2}$$

The modeling of hydropower units in OpenHPL is considered using $\dot{m}$ as *a flow variable* and pressure p at any point in a unit as *a across variable*[8].

Figure 2 shows a connector for a hydropower unit created for OpenHPL. While developing a unit, for example a surge tank in our case, mathematical terms in mass and momentum balances as expressed in Eq. 1 and Eq. 2 should be reduced using algebraic variables to $\dot{m}$, p_b, and p_t. For the surge tank we have $p_i = p_b$ and $p_o = p_t$ for a volumetric flow $\dot{V}$ sign convention to be positive in the upward direction as in Fig. 1.

The mass balance equation represented in Eq. 1 can be represented by series of algebraic equations. The mass of water inside the surge tank leads to,

$$m = \rho A \ell \tag{3}$$

$$\ell = \frac{h}{\cos\theta} \tag{4}$$

$$A = \frac{\pi D^2}{4}, \tag{5}$$

and the mass flow rate leads to,

$$\dot{m} = \rho \dot{V}, \tag{6}$$

where A and θ represents cross-sectional area and slant angle of a cylindrical surge tank with diameter D. ρ represents density of the fluid. h and $\dot{V}$ are differential variable with initial height of water level inside the surge tank as h_o and initial discharge to the surge tank as $\dot{V}_o$. Similarly, the momentum balance equation in Eq. 2 can be represented by a series of algebraic equations as,

$$\mathcal{M} = mv \tag{7}$$

$$v = \frac{\dot{V}}{A} \tag{8}$$

$$\dot{\mathcal{M}} = \dot{m}v \tag{9}$$

$$F = F_p - F_f - F_g, \tag{10}$$

---

[8]The across and flow variables are used for creating a connectors in Modelica language. For more details, please see https://mbe.modelica.university/components/connectors/ .

where v is the velocity of water inside the surge tank in the direction of $\dot{V}$ and F is the total force acting inside the surge tank in the direction of v. F_p is the force exerted due to pressure difference $p_b - p_t$. F_f is the fluid friction acting opposite in the direction of v, and F_g is directed downward due to gravity.

The pressure force exerted in the normal direction of A is given by,

$$F_p = (p_b - p_t)A. \tag{11}$$

Furthermore, the fluid friction F_f is calculated as,

$$F_f = \frac{K''' A_w f_D}{4} \tag{12}$$

$$A_w = \pi D \ell \tag{13}$$

$$K''' = \frac{\rho \dot{V} \mid \dot{V} \mid}{2A^2}, \tag{14}$$

where K''' is the kinetic energy of the fluid per volume which is proportional to the quadratic variation of $\dot{V}$ and v. The expression for K''' is $\frac{\rho v |v|}{2}$. The absolute value for v and $\dot{V}$ allow for reversing direction of water flow. A_w represent the wetted area due to water flow inside the surge tank given by an expression $A_w = \pi D \ell$. In Eq. 12, f_D represents Darcy's friction factor given by an *implicit* expression in the Colebrook–White equation (Colebrook and White, 1937; Colebrook et al., 1939) for *transient* full-fluid flow in the conduit. There exists several *explicit* approximation for f_D that requires less computation as listed in (Lie, 2017a, p. 239). For OpenHPL, we are using the explicit approximation of Colebrook–White equation from (Swanee and Jain, 1976),

$$\frac{1}{\sqrt{f_D}} = -2 \log_{10} \left(\frac{\varepsilon/D}{3.7} + \frac{5.7}{N_{Re}^{0.9}} \right), \tag{15}$$

for $N_{Re} = \left(2300 - 10^8 \right)$ and $\varepsilon/D = \left(10^{-5} - 0.005 \right)$, where ε is a conduit roughness height and N_{Re} is the Reynolds number expressed by $N_{Re} = \frac{\rho |v| D}{\mu}$. Here, μ represents kinematic viscosity of the fluid.

For *laminar* flow, $f_D = \frac{64}{N_{Re}}$ with $N_{Re} < 2100$. The region for the fluid with $2100 \le N_{Re} < 2300$ is a *transition* flow interpolated with a 4th order polynomial equation.

The expression for force due to gravity is given as,

$$F_g = mg \cos \theta. \tag{16}$$

Equation 1 to 16 represents Differential Algebraic Equations (DAEs) for the mechanistic modeling of simple surge tank represented in Fig. 1 and can be solved using equation based modeling language like Modelica.

2.1.2 Sharp orifice type surge tank

The model of the simple surge tank can be further modified using a sharp orifice hydraulic obstruction inside the surge tank as represented in Fig. 3.

To model the sharp orifice type surge tank we employ the *generalized friction factor* for sharp orifice fitting as
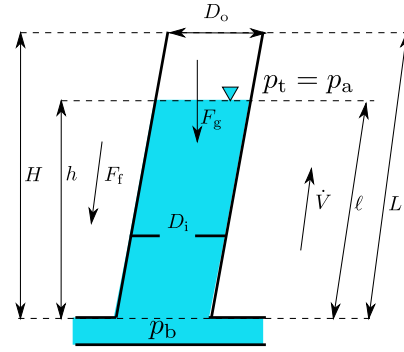
Figure 3. Sharp orifice type surge tank with a sharp orifice of diameter D_i as shown by horizontal perturbation bars inside the surge tank. The sharp orifice acts as a hydraulic obstruction for water flowing inside the surge tank.

given in (Lie, 2017a, p. 246). The expression for fluid friction force represented by Eq. 12 needs a correction term due to the sharp orifice. The frictional force exerted due to sharp orifice can be calculated using an expression for a pressure drop expression as given in (Lie, 2017a, p. 244). The overall frictional force for the sharp orifice type surge tank is now calculated by the expression as,

$$F_f = \frac{K''' A_w f_D}{4} + \frac{1}{2} \rho v \mid v \mid A \phi_{so}, \tag{17}$$

where A is the cross-sectional area of the sharp orifice type surge tank with diameter D_o which is equivalent to the simple surge tank with diameter noted with symbol D, and ϕ_{so} is a generalized friction factor. ϕ_{so} depends on N_{Re}, and the diameter of the surge tank and the orifice.
For $N_{Re} < 2500$:

$$\phi_{so} = \left[2.72 + \left(\frac{D_i}{D_o} \right)^2 \left(\frac{120}{N_{Re}} - 1 \right) \right] \cdot \phi_{so}^0$$

For $N_{Re} \ge 2500$:

$$\phi_{so} = \left[2.72 + \left(\frac{D_i}{D_o} \right)^2 \cdot \frac{4000}{N_{Re}} \right] \cdot \phi_{so}^0$$

where,

$$\phi_{so}^0 = \left[1 - \left(\frac{D_i}{D_o} \right)^2 \right] \left[\left(\frac{D_i}{D_o} \right)^4 - 1 \right].$$

Equation 1 to 11, Eq.17, and Eq.13 to 16 represent DAEs for the mechanistic model of the sharp orifice type surge tank.

2.1.3 Throttle valve surge tank

To model a throttle valve surge tank we employ a similar method for correction of fluid frictional force as in the case of a sharp orifice type surge tank. A schematic diagram for throttle valve surge tank is given in Fig. 4.
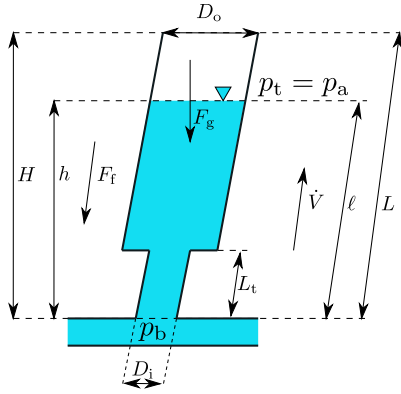
Figure 4. Throttle valve surge tank with the diameter and the length of throat as D_i and L_t, respectively. The throat in the figure acts as a hydraulic obstruction for the water flow inside the surge tank.

The water mass, velocity, momentum, and the fluid friction force for a throttle valve surge tank varies depending upon the water level inside the surge tank as above or below the throat.

1. **For $\ell \leq L_t$:** When the water level is at the throat or below the throat, we have,

$$m = \rho A_t \ell \tag{18}$$

$$F_f = \frac{K''' A_{w,t} f_D}{4}, \tag{19}$$

where A_t is the area of throat, $A_{w,t}$ is wetted area for the throat given by expression $A_{w,t} = \pi D_t \ell$. The average velocity for this case is calculated using $v = \frac{\dot{V}}{A_t}$ and K''' is given by expression $K''' = \frac{1}{2}\rho v \mid v \mid$

2. **For $\ell > L_t$:** When water level inside the surge tank is above the throat of the surge tank, the frictional factor due to the throat should be considered.

The mass of the water inside the surge tank in this case is given as,

$$m = \rho(A_t L_t + A(\ell - L_t)), \tag{20}$$

where $A = \frac{\pi D_o^2}{4}$. Similarly, the average velocity, water momentum, and pressure force are given as,

$$v = \frac{\dot{V}}{2}\left(\frac{1}{A_t} + \frac{1}{A}\right)$$
$$\mathcal{M} = \rho \dot{V} \ell$$
$$F_p = (p_b - (p_t + \rho g(\ell - L_t)))A_t + \rho g(\ell - L_t)A.$$

Observe that while considering the frictional force correction factor for $\ell > L_t$ the velocity direction is important.

a). **For $v \geq 0$:** The friction factor is calculated considering a *square expansion* type pipe fitting and the generalized friction factor is given as (Lie, 2017a, p. 245)
For $N_{Re} < 4000$:

$$\phi_{se} = 2\left[1 - \left(\frac{D_o}{D_i}\right)^4\right]$$

For $N_{Re} \geq 4000$:

$$\phi_{se} = (1 + 0.8 f_D)\left[1 - \left(\frac{D_o}{D_i}\right)^2\right].2$$

This gives the total frictional force for this case,

$$F_f = \frac{K''' A_w f_D}{4} + \frac{1}{2}\rho v_e \mid v_e \mid A_t \phi_{se}, \tag{21}$$

where ϕ_{se} represents the generalized friction factor for the square expansion type fitting. The *entrance velocity* for square expansion type fitting is expressed as $v_e = \frac{\dot{V}}{A_t}$ and the *entrance area* is A_t. The wetted area is calculated using $A_w = \pi D(\ell - L_t)$.

b). **For $v < 0$:** The flow of water in this case is considered to be from the top of the surge tank to the bottom direction. The friction factor is calculated considering *square reduction* type pipe fitting and the generalized friction factor is given as,
for $N_{Re} < 2500$:

$$\phi_{sr} = \left(1.2 + \frac{160}{N_{Re}}\right)\left[\left(\frac{D_i}{D_o}\right)^4 - 1\right]$$

and for $N_{Re} \geq 2500$:

$$\phi_{sr} = (0.6 + 0.48 f_D)\left(\frac{D_i}{D_o}\right)^2\left[\left(\frac{D_i}{D_o}\right)^2 - 1\right].$$

This gives the total frictional force for this case to be,

$$F_f = \frac{K''' A_w f_D}{4} + \frac{1}{2}\rho v_e \mid v_e \mid A \phi_{sr}, \tag{22}$$

where ϕ_{sr} represents the generalized friction factor for the square reduction type fitting. The *entrance velocity* is expressed as $v_e = \frac{\dot{V}}{A}$ and the *entrance area* is A.

2.1.4 Air-cushion surge tank

The general schematic of air-cushion surge tank is shown in Fig. 5. The free water surface inside the surge tank is filled with pressurized air making it as a closed type surge tank.

The pressure wave during *a load rejection* traveled from *high pressure region* (at the end of penstock) to the *low pressure region* (near free water surface, i.e., through the surge tank in hydro power systems). During this period, water mass inside the surge tank oscillates, dissipating pressure. The more the amplitude of water mass oscillation the higher should be the physical height of the surge tank. For reducing the amplitude of water oscillation inside the surge tank, pressurized air is placed inside the surge tank making a closed surge tank. This will cause the air to compress and expand adiabatically (Vereide et al., 2016; Zhang et al., 2009), and the energy due to high pressure is realeased as a form of work done for compression and rarefaction. For a adiabatic process we have,

$$pV^\gamma = \text{constant}, \tag{23}$$
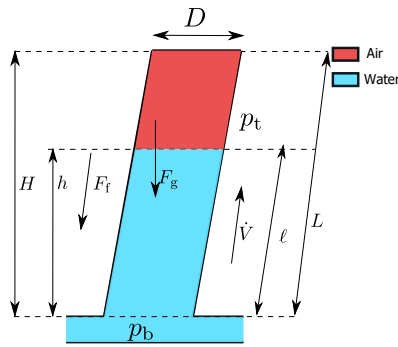
Figure 5. Air-cushion surge tank with initial air pressure of p_t which is normally tens of atmospheric pressure.

Figure 6. Conical diffuser inclined at angle θ with input diameter D_i and output diameter D_o. p_i and p_o are input and output pressure of the conical diffuser with $p_i > p_o$.

where p, V, and γ is the pressure, volume, and ratio of specific heats at constant pressure and at constant volume, respetively, for air. The mechanistic model of air cushion surge tank is similar to that of simple surge tank, however with correction term for m and p_t for a simple surge tank.

The mass of water and air inside surge tank is given as,

$$m = \rho A \ell + m_a, \tag{24}$$

and the air pressure inside the surge tank is given by,

$$p_t = p_c \left(\frac{L - \ell_o}{L - \ell} \right)^\gamma, \tag{25}$$

where m_a is the mass of air inside the surge tank given by expression,

$$m_a = \frac{p_c A (L - \ell_o) M_a}{R T^o}. \tag{26}$$

In Eq. 25, p_c is the initial air cushion pressure when initial slant height of liquid level inside the surge tank is ℓ_o. The expression shown in Eq. 25 is derive from Eq. 23 equalizing the *initial* and *final* expression. In Eq. 26, M_a represents molar mass of air, R is the universal gas constant, and T^o is the temperature of air inside the surge tank.

2.2 Draft tube

A draft tube is a hydraulic device used in reaction turbines in a hydro power systems for utilizing the available kinetic energy at the exit of the runner of the turbine. One of the prime benefits of a draft tube is to increase the turbine pressure head by decreasing the exit velocity out of the runner which will improves the overall efficiency of hydropower systems. And the other benefit is that the back flow of water from *a tailrace* to the turbine is restricted due to higher pressure region at the turbine's outlet due to draft tube (Gubin, 1973). There are various types of draft tubes; the most common type is a conical diffuser and others are variants of the conical diffuser (Arasu, 2008). In this paper, we will derive a mechanistic model for a conical diffuser and a hydraucone or a Moody spreading pipes (White, 1921).

2.2.1 Conical diffuser

A general schematic of the conical diffuser is shown in Fig. 6.

The influent and effluent mass flow of water through the conical diffuser is same. This gives $\frac{dm}{dt} = 0$ from Eq. 1. Thus, the mechanistic model is derived from the momentum balance given by Eq. 2 with series of DAEs. First, we consider the model of the conical diffuser considering it be a cylinder of average diameter $D = \frac{D_i + D_o}{2}$. Second, we will add a frictional force correction factor for the conical diffuser expanded from D_i to D_o (with *a diffusion angle* normally in the range of $(5° - 20°)$).

$$\frac{d\mathcal{M}}{dt} = \dot{\mathcal{M}} + F \tag{27}$$

$$M = mv \tag{28}$$

$$v = \frac{\dot{V}}{A} \tag{29}$$

$$\dot{\mathcal{M}} = \dot{m}v \tag{30}$$

$$F = F_p - F_g - F_f, \tag{31}$$

where $F_p = p_i A_i - p_o A_o$, $A_i = \frac{\pi D_i^2}{4}$, $A_o = \frac{\pi D_o^2}{4}$, $A = \frac{\pi D^2}{4}$, and $F_g = mg \cos \theta$. The mass of water inside the diffuser is given by,

$$m = \rho V,$$

where V is the volume of water. The expression for V can be calculated as[9],

$$V = \frac{\pi H}{12} \left(D_i^2 + D_o^2 + D_i D_o \right).$$

The overall frictional force is calculated using expression

$$F_f = \frac{K''' A_w f_D}{4} + \frac{1}{2} \rho v \mid v \mid A_i \phi_d, \tag{32}$$

---

[9]https://mathworld.wolfram.com/ConicalFrustum.html
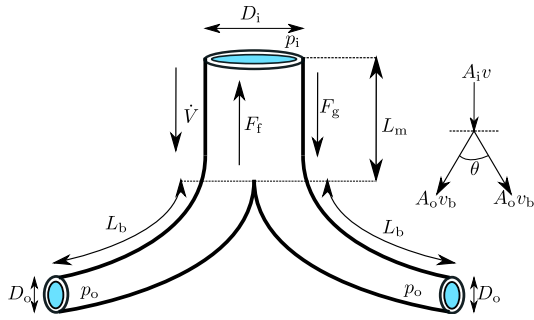
Figure 7. Moody spreading pipes with length of *main part* L_m with both the *branch* length of L_b. v and v_b are velocity through the main part and the branch part, respectively. The continuity equation for the pipe branching is $A_i v = A_o v_b + A_o v_b$. The pipe contracts from the point of branching to the outlet of the branch in the real case of hydraucone as explain in (White, 1921, p. 276). However, we are considering a constant cross-section throughout the branch pipe. θ is a branching angle or a bifurcation angle.

where ϕ_d is the generalized friction factor due to the diffusion.

The *head loss* for a conical diffuser, diffused from D_i to D_o, is minimum at a diffusion angle of $8°$ for a fixed value of $\frac{D_o}{D_i}$. For *a pair value* of diffusion angle and the ratio $\frac{D_o}{D_i}$, ϕ_d can be calculated from (Munson et al., 2009, p. 420). For our case, for a maximum efficiency conical diffuser, we will conside a diffusion angle of $8°$ which gives

$$\phi_d \approx 0.23 \left(1 - \frac{D_i}{D_o} \right)^2. \tag{33}$$

The mechanistic model of a conical diffuser can be represented by using DAEs from Eq. 27 to 33.

2.2.2 Moody spreading pipes

The schematic diagram of a moody spreading pipes or a hydraucone is shown in Fig. 7.

For Moody spreading pipes, $\frac{dm}{dt} = 0$, and the mechanistic model is developed from the momemtum balance. We take the momentum balance considering verticle direction i.e., $y - axis$ momentum conservation. The series of DAEs are,

$$\frac{d\mathcal{M}}{dt} = \dot{\mathcal{M}} + F \tag{34}$$

$$\mathcal{M} = m_m v_m + 2 m_b v_b \cos \frac{\theta}{2} \tag{35}$$

$$\dot{\mathcal{M}} = \dot{m}_m v_m + 2 \dot{m}_b \cos \frac{\theta}{2} \tag{36}$$

$$F = F_p - F_f - F_g, \tag{37}$$

where m_m and m_b are mass of water in the main part and the branching part, respectively given by expressions $m_m = \rho A_i L_m$ and $m_b = \rho A_o L_b$. Similarly, v_m and v_b are the velocity in the main and the branching part, respectively. $v_m = \frac{\dot{V}}{A_i}$ and v_b are calculated using continuity equation

Table 1. ϕ_d^o for different value of θ for Moody spreading pipes friction factor correction

θ	15	30	45	60	90
ϕ_d^o	0.04	0.16	0.36	0.64	1

for branching pipes using expression as,

$$A_i v_m = A_o v_b + A_o v_b$$
$$A_i v_m = 2 A_o v_b$$
$$v_b = \frac{A_i}{2 A_o} v_m.$$

Furthermore, expressions for $\dot{m}_m$ and $\dot{m}_b$ are given as,

$$\dot{m}_m = \rho \dot{V}$$
$$\dot{m}_b = \rho \dot{V}_b$$
$$\dot{V}_b = A_o v_b,$$

where $\dot{V}_b$ is the volumetric flow rate in the brach.

The components in Eq. 37 are expressed as,

$$F_p = p_i A_i - 2 p_o A_o \cos \frac{\theta}{2} \tag{38}$$

$$F_g = m_m g + 2 m_b g \cos \frac{\theta}{2}. \tag{39}$$

The fluid frictional force F_f is calculated considering fluid friction in the main and the brach pipe with addition of a generalized frictional force correction factor for branching. The overall frictional force is then,

$$F_f = F_{f,m} + 2 F_{f,b} \cos \frac{\theta}{2} + 2 \cdot \frac{1}{2} \rho v_m \mid v_m \mid A_i \phi_d, \tag{40}$$

where ϕ_d is the generalized friction factor for a single branch in case of branching and taken from (Idelcik, 1966, p. 281, 301), given as,

$$\phi_d = 1 + \left(\frac{v_b}{v_m} \right)^2 - 2 \frac{v_b}{v_m} \cos \theta - \phi_d^o \left(\frac{v_b}{v_m} \right)^2, \tag{41}$$

where ϕ_d^o depends on θ and calculated from Table 1.

$F_{f,m}$ and $F_{f,b}$ are fluid frictional force, calculated using Eq. 12, for main and the branch pipe for Moody spreading pipes.

Equation 34 to 41 represent DAEs for mechanistic modeling of Moody spreading pipes or Hydraucone.

3 Simulated Responses and Results

For the simulated responses from the mechanistic models of surge tank and draft tubes we take a case study from a real hydropower plant. For simulating open surge tanks and draft tubes we are using the layout of the Trollheim hydro power plant and for simulating air-cushion surge tank we are using the layout of the Torpa hydropower plant.

The layout diagram of Trollheim and Torpa hydro power plants are shown in Fig. 8.
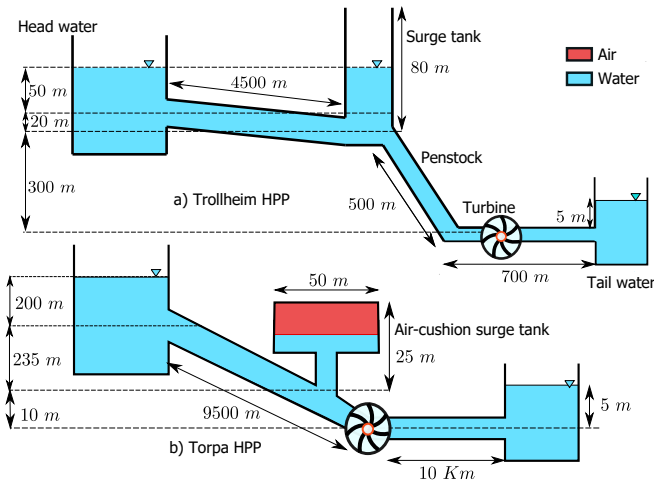
Figure 8. Layout diagram for the Trollheim (Vytvytskyi and Lie, 2019a) and the Torpa Hydro Power Plant (HPP) (Vereide et al., 2016). Nominal head, nominal discharge, and nominal power output are 370 m, 40 m^3/s and 130 MW for the Trollheim HPP, and 445 m, 35 m^3/s and 150 MW for Torpa HPP. Torpa HPP has two turbine units each having nominal power output of 75 MW. The air-cushion surge tank for the Torpa HPP has air volume of 13,000 m^3 initially pressurized at 4.1 Mpa. For the Trollheim HPP, the diameter for both of the penstock and the surge tank is 4 m while for both of the headrace and the tailrace tunnel is 6 m. Similarly, for the Torpa HPP, the diameter of both of the headrace and the tailrace tunnel is 7 m.

3.1 Responses for surge tanks

Figure 9 shows the simulated response for different surge tanks for the Trollheim HPP and the Torpa HPP.

In case of the Trollheim HPP, the manifold pressure surge during *load acceptance*[10] in case of a simple surge tank has higher amplitude than that for sharp orifice and throttle valve surge tank. Furthermore, the pressure surge dies out soon in case of both sharp orifice and throttle valve surge tank. Also, the diameter of sharp orifice and throttle valve affects the manifold pressure surge oscillation. For example, when $D_t = 1$ m for a throttle valve surge tank the manifold pressure p_b settles after 20 s while for sharp orifice type with $D_{so} = 1$ m, p_t settles around 300 s.

In case of the Torpa HPP, the dynamics of both manifold pressure and air-cushion pressure is the same with a difference of almost 2 bars.

3.2 Responses for draft tubes

Figure 9 shows the simulated response for a conical diffuser and Moody spreading pipes for the the Trollheim HPP.

For a Moody spreading pipes draft tube, the inlet pressure p_i decreases as the branching angle decreases.

---

[10]It is the condition when the load at prime mover is added suddenly. For instance, in case of turbine running an electrical generator the sudden industrial operation like electrical motors, etc., can be considered as a load acceptance.
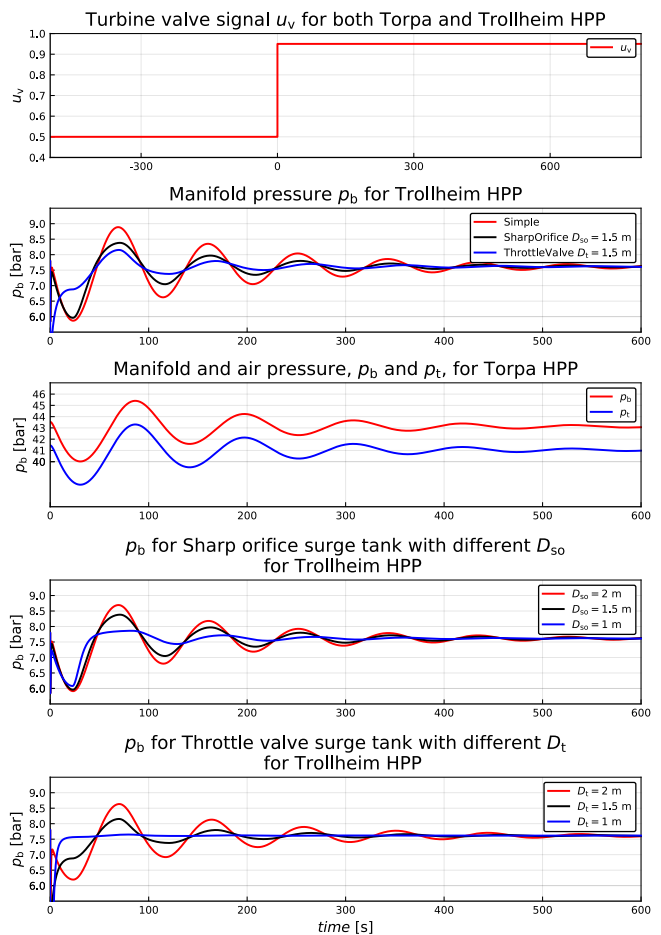
Figure 9. Simulated response for different surge tanks with step change of 0.45 to turbine's valve signal. The turbine's valve signal starts at $t = -500$ s to show that simulation is performed in steady-state for -500 s and at $t = 0$ s the valve signal is change from 0.5 to 0.95 to see the dynamics of other variables.

4 Conclusions and Future Work

This paper consists of mechanistic models for different types of surge tanks and draft tubes. Result shows that the pressure surge during load acceptance dies out soon for the throttle valve surge tank when the diameter of throat is decreased succesively. Similar operations can be obtained for a sharp orifice type surge tank, however, pressure surge amplitude decays soon in case of a throttle valve surge tank. For an air-cushion surge tank, the dynamics of both manifold pressure and air-cushion pressure are the same. For a Moody spreading pipes, the inlet pressure decreases when the braching angle is decreased.

Future work includes testing of the surge tank and draft tube mechanistic models with experimental data. The dynamics of sharp orifice type surge tank and throttle valve surge tank can be validated with simulated results from (Bhattarai et al., 2019). The air-cushion surge tank model can be validated with experimental results obtained from (Vereide et al., 2016). Similarly, model validation for the conical diffusers can be done with experimental data from
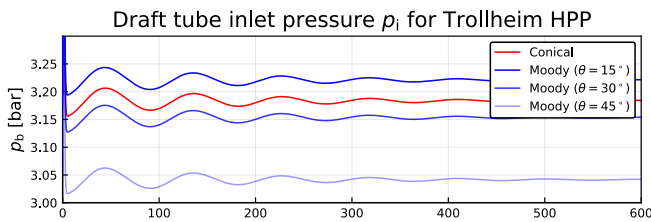
Figure 10. Simulated response for conical diffuser and Moody spreading pipes for different branching angle. $D_i = 4\,\mathrm{m}$ for both types of draft tubes. $D_o = 4.978\,\mathrm{m}$ for conical diffuser with diffusion angle of $8°$ and $D_o = 3.5\,\mathrm{m}$ for Moody spreading pipes. $L_m = 4\,\mathrm{m}$ and $L_b = 3\,\mathrm{m}$ for Moody spreading pipes. Moody sprading pipes draft tube is simulated for different branching angle of $15°$, $30°$, and $45°$.

(Vytvytskyi and Lie, 2019a). For Moody spreading pipes, prior model judgement is required from experts.

References

A Valan Arasu. *Turbo Machines*. Vikas Publishing House, 2008.

Khem Prasad Bhattarai, Jianxu Zhou, Sunit Palikhe, Kamal Prasad Pandey, and Naresh Suwal. Numerical modeling and hydraulic optimization of a surge tank using particle swarm optimization. *Water*, 11(4):715, 2019.

R. Lester Brown. Plan B Updates - 107: The Great Transition, Part I: From Fossil Fuels to Renewable Energy | EPI, Aug 2012. http://www.earth-policy.org/plan_b_updates/2012/update107.

CF Colebrook and CM White. Experiments with fluid friction in roughened pipes. *Proceedings of the Royal Society of London. Series A-Mathematical and Physical Sciences*, 161 (906):367–381, 1937.

Cyril Frank Colebrook, T Blench, H Chatley, EH Essex, JR Finniecome, G Lacey, J Williamson, and GG Macdonald. Correspondence. turbulent flow in pipes, with particular reference to the transition region between the smooth and rough pipe laws.(includes plates). *Journal of the Institution of Civil engineers*, 12(8):393–422, 1939.

Maksimillian Fedorovich Gubin. *Draft tubes of hydro-electric stations*. Published for the Bureau of Reclamation, US Dept. of the Interior and âŠ, 1973.

IE Idelcik. Handbook of hydraulic resistance: coefficients of local resistance and of friction. Technical report, ERDA Div. Phys. Res., 1966.

Bernt Lie. Modeling of Dynamic Systems, August 2017a. Lecture notes, Version of August 8.

BR Munson, DF Young, and TH Okiishi. Fundamentals of Fluid Mechanics, 2009.

Madhusudhan Pandey and Bernt Lie. The Role of Hydropower Simulation in Smart Energy Systems. 2020.

Valentyna Splavska, Liubomyr Vytvytskyi, and Bernt Lie. Hydropower systems: comparison of mechanistic and table

look-up turbine models. In *Proceedings of the 58th Conference on Simulation and Modelling (SIMS 58) Reykjavik, Iceland, September 25th–27th, 2017*, number 138, pages 368–373. Linköping University Electronic Press, 2017.

PK Swanee and Akalank K Jain. Explicit equations for pipeflow problems. *Journal of the hydraulics division*, 102(5), 1976.

Kaspar Vereide, Bjørnar Svingen, Torbjørn Kristian Nielsen, and Leif Lia. The effect of surge tank throttling on governor stability, power control, and hydraulic transients in hydropower plants. *IEEE Transactions on Energy Conversion*, 32(1):91–98, 2016.

Liubomyr Vytvytsky and Bernt Lie. Comparison of elastic vs. inelastic penstock model using OpenModelica. In *Proceedings of the 58th Conference on Simulation and Modelling (SIMS 58) Reykjavik, Iceland, September 25th–27th, 2017*, number 138, pages 20–28. Linköping University Electronic Press, 2017. doi:http://dx.doi.org/10.3384/ecp1713820.

Liubomyr Vytvytskyi. *Dynamics and model analysis of hydropower systems*. PhD thesis, University of South-Eastern Norway, 2019. URL http://hdl.handle.net/11250/2608105.

Liubomyr Vytvytskyi and Bernt Lie. Mechanistic model for Francis turbines in OpenModelica. *IFAC-PapersOnLine*, 51(2):103–108, 2018. doi:https://doi.org/10.1016/j.ifacol.2018.03.018.

Liubomyr Vytvytskyi and Bernt Lie. OpenHPL for Modelling the Trollheim Hydropower Plant. *Energies*, 12(12):2303, 2019a.

Liubomyr Vytvytskyi and Bernt Lie. Combining measurements with models for superior information in hydropower plants. *Flow Measurement and Instrumentation*, 69:101582, 2019b.

William M White. *The Hydraucone Regainer, Its Development and Applications in Hydro-electric Plants*. 1921.

Wikipedia. Nuclear power debate, Oct 2019. https://en.wikipedia.org/wiki/Nuclear_power_debate.

Xiaohong Zhang, Zhengfei Zhao, and Yanhang Zhang. Study on wave calculation of an air cushion surge chamber. In *2009 International e-Conference on Advanced Science and Technology*, pages 70–73. IEEE, 2009.

Semi-Automatic Optimization of Steel Heat Treatments for Achieving Desired Microstructure

Aarne Pohjonen[1]   Vahid Javaheri[1]   Joni Paananen[1]   Juha Pyykkönen[2]

[1]Materials and Mechanical Engineering, University of Oulu, Finland, {Aarne.Pohjonen, Vahid.Javaheri, Joni.Paananen}@Oulu.fi
[2]SSAB Europe, Finland, Juha.Pyykkonen@SSAB.com

Abstract

The thermo-mechanical processing history together with the steel composition defines the final microstructure, which in turn produces the macroscopic mechanical properties of the final product. In many industrial processes it is therefore of paramount importance to find the optimal thermal path that produces the desired microstructure. In the current study an optimization method has been developed to calculate the optimal thermal path for producing desired amounts of microstructural constituents (ferrite, bainite, martensite) of a medium carbon, low-alloy steel, and a low carbon microalloyed steel. The optimization is performed for two separate industrial processes: induction hardening of a pipeline steel and a water cooling of hot rolled steel strip. The optimization workflow consists of first setting the desired amounts of microstructural constituents, and subsequent optimization of the thermal path, which produces these desired amounts. For the water cooling of a steel strip we additionally employed previously developed tool to calculate the cooling water fluxes that are needed to realize the optimized cooling path in water cooling line after hot rolling. To demonstrate the applicability of the method, we present results that were obtained for different case studies related to the industrial processes.

Keywords: constrained optimization, nonlinear optimization, steel, processing

1   Introduction

1.1   Background

The thermo-mechanical processing history together with the steel composition defines the final microstructure, which in turn produces the macroscopic mechanical properties of the final product. In many industrial processes, it is therefore of paramount importance to find the optimal thermal path that produces the desired microstructure.

In the case of pipe induction hardening, the gradient of material properties (microstructure, hardness and hardening depth) through the pipe body can be optimized by designing the thermal cycle i.e. rates of heating and cooling along with the peak temperature and soaking time. In the induction hardening

processing, the thermal cycle is very easy and precise to apply and control. As regards, an optimization is required to reach the desired material properties through the final microstructure characteristics in an efficient way.

Higher heating rate is beneficial in many ways. However to some extent it can cause inhomogeneity and abnormal grain structure (Javaheri *et al.*, 2019a) in the steel. The heating rate can also change the critical phase transformation temperature, which cause difficulties for selecting the austenitization temperature. The minimum possible peak temperature can result in the finest prior austenite grain structure, which leads to formation of very fine final microstructure and consequently will improve the strength and toughness at the same time. Increasing the cooling rate in the industrial scale may affect the production cost. However producing hard and strong microstructure such as martensite or lower bainite would not be possible if the cooling rate is smaller than a certain level. All in all, an optimization should be performed to find the best and most efficient combination of above mentioned parameters.

In industrial practice of hot rolling of steel, steel manufacturers are forced to employ the most advanced processing routes in their production lines. The control of reheating, rolling and accelerated cooling forms the basis of producing new types of high-performance steels. The capability to accurately predict and control plate or strip temperature behavior and microstructure evolution during thermomechanical processing is of fundamental importance in commercial production. In the water cooling line, it is possible, within certain limits, control the resulting microstructure for a given steel composition. Since the line layout and other processing parameters limit the available cooling paths, the optimization script needs to take these effects in to account conveniently.

Optimization of thermal processing is a common task arising in metal processing (Jung *et al.*, 2018; Tavakoli, 2018). In current study we employ the optimization tools in Matlab programming language.

1.2   Aims

In the current study an optimization tool has been developed to calculate the optimal thermal path for

producing desired amounts of microstructural constituents (e.g. ferrite, bainite, martensite). Since in the current implementation, the minimization algorithm applies an initial guess, we call this implementation as "semi-automatic", although for the water cooling of steel strip the choice of the initial guess was performed by automatically testing different cooling paths, as explained later.

The optimization is performed for two separate industrial processes: induction hardening of a medium carbon, low-alloy pipe steel (Javaheri *et al.*, 2019b) and a water cooling of a hot rolled low carbon steel strip (Pohjonen *et al.*, 2018a). The optimization workflow consists of first setting the desired amounts of microstructural constituents, and subsequent optimization of the thermal path, which produces these desired amounts applying the phase transformation model described in (Pohjonen *et al.*, 2018a; Pohjonen *et al.*, 2018b; Javaheri *et al.*, 2019b). For the water cooling of a steel strip, we additionally employed previously developed tool (Paananen, 2015; Pohjonen *et al.*, 2016) to calculate the cooling water fluxes that are needed to realize the optimized cooling path in water cooling line after hot rolling.

To optimize the induction hardening process for achieving desired gradient of microstructure and mechanical properties, initially, the heating cycle (austenitization) has been optimized regarding the rate of heating and peak temperature assuming a dwell time of 2.5 s aiming for the finest prior austenite structure. Then, three different scenarios for the cooling path have been considered to achieve three thoroughly different microstructures of i) fully martensitic structure (100%), ii) fully bainitic microstructure, equal amount of upper and lower bainite (50% each) and iii) mixture of martensite and bainite (30% martensite, 40% upper bainite, and 30% lower bainite).

To test the optimization tool in the context of water cooling of hot rolled steel strip, we set the desired fractions of bainite and martensite, and perform optimization to achieve these fractions within the water cooling line. In the case of water cooling, the line layout and rolling process set some limits for the possible cooling rates. Maximum overall cooling rate at the strip centerline ranges typically from 150 °C/s to 30 °C/s while minimum cooling rate due to radiation and convection ranges from 10 °C/s to 2 °C/s for the thin and thick plates, respectively.

2 Calculations

In the current study, an optimized piecewise linear cooling path was sought. The optimization of the path was realized with the Matlab Optimization Toolbox (Mathworks, 2020) using the fmincon function. The interior point minimization algorithm with nonlinear constraints (Byrd *et al.*, 2000) was chosen for the optimization. Since the heating and cooling rates are limited in practice, the constraints for these rates are defined for the optimization. Also, the maximum temperature changes and time intervals of the path can be limited to desired values. By setting the exact time intervals for the path, the optimization tool can be adjusted to given water cooling line layout, when the speed of the strip is known.

The workflow for the cooling path optimization is depicted in Fig. 1. The applied optimization method requires an initial guess for the optimized parameters, which is used as a starting point for the optimization. Here we use the starting temperature T_0, the temperature changes ΔT_i and time intervals Δt_i for each segment i of the piecewise linear temperature path as the optimization parameters, so that the parameters describe the thermal path. The thermal path optimizer applies the fmincon function, which runs a script that calculates the phase fractions that result from the thermal path. After the calculation, the optimizer script receives the final phase fractions, and calculates the difference to the desired phase fractions, $_i \chi_{i,des} - \chi_i$. Based on the optimization algorithm (Byrd et al., 2000; Mathworks, 2020) the thermal path is changed and the optimization proceeds to minimize the difference to the desired phase fractions.
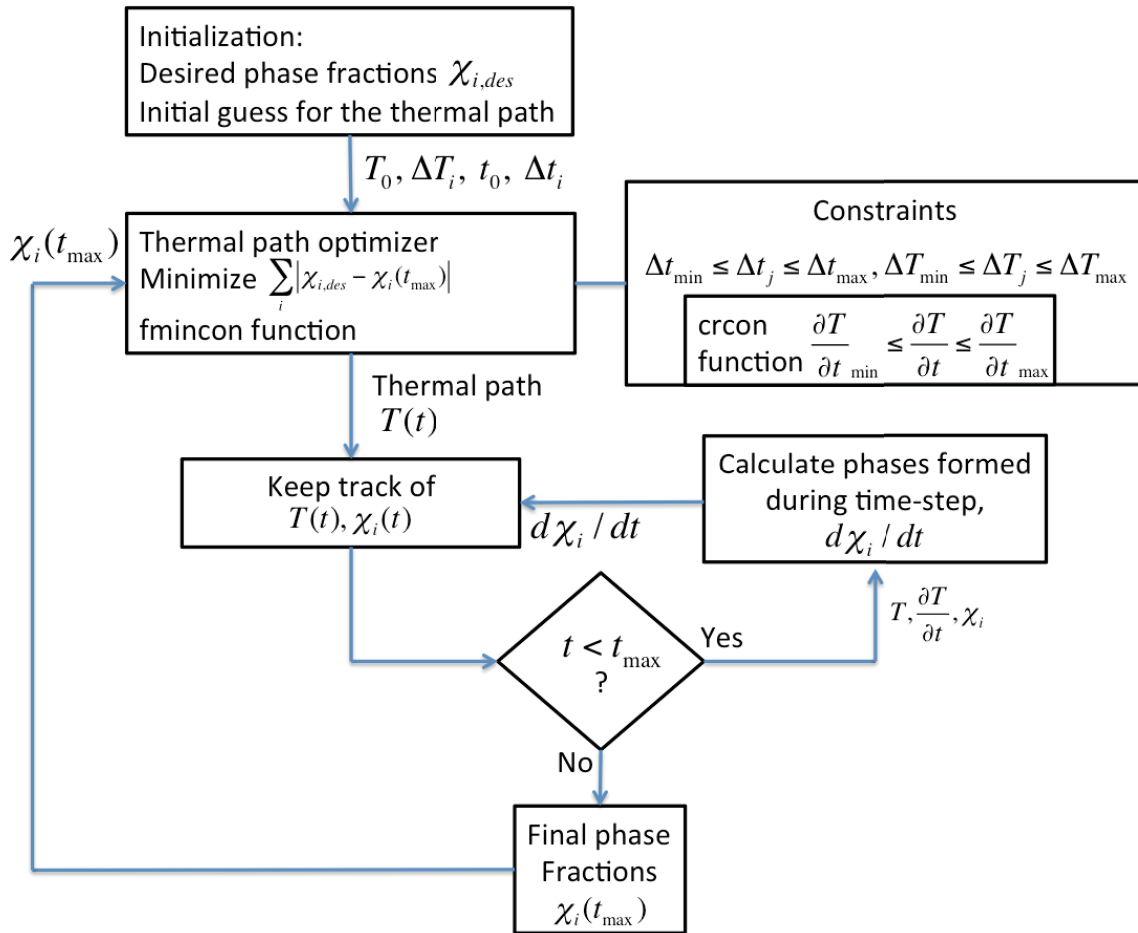
Figure 1. Optimization flowchart

Upper and lower bounds were set within the fmincon function call. From the application point of view, it is important to define the minimum and maximum cooling and heating rates, since even without additional cooling, the object cools by itself due to radiation and convection. Also, depending on the size and geometry and the applied forced cooling, there is maximum cooling rate that can be achieved. Too high or low heating rates can adversely affect the resulting microstructure. For this reason, the non-linear constraints for heating and cooling rates $\partial T/\partial t$ were set using crcon function (Mathworks, 2020) related to the fmincon minimization algorithm, to control the maximum cooling rate for each segment in the piecewise linear cooling path: $\frac{\Delta T}{\Delta t}_{min} \leq \frac{\Delta T_i}{\Delta t_i} \leq \frac{\Delta T}{\Delta}_{max}$. The phase transformation module has been described in detail generally in (Pohjonen *et al.*, 2018b), in context of induction hardening of medium carbon steel (Javaheri *et al.*, 2019b) and in the context of water cooling of hot rolled low carbon steel in (Pohjonen *et al.*, 2018a), where (Javaheri *et al.*, 2019b) and (Pohjonen *et al.*, 2018a) also include the transformation model parameters fitted for the steels that were used in these examples.

Since the minimization script finds a local minimum near to the initial guess, it is required to choose suitable initial guess, so that the algorithm can find a good minimum. For the test cases related to optimization of the cooling path of water cooled steel strips, we applied a script to automatically test nine different initial guesses for the cooling paths and then select the one that produced the best solution. The cooling paths described by the initial guesses are shown in Fig. 2.
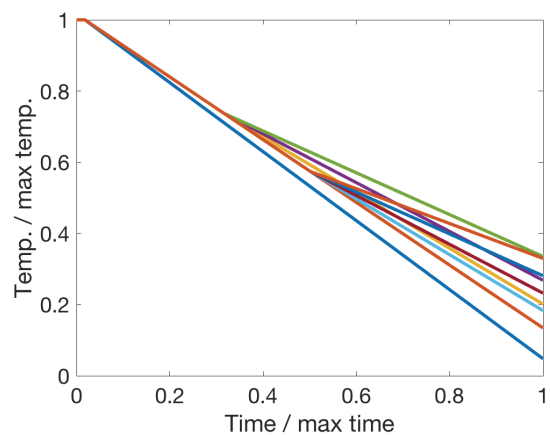
Figure 2. Thermal paths used as initial guesses for the optimization of the temperature path for water cooling of hot rolled steel strip.

After the thermal path to produce desired microstructure in the strip cooling line is optimized, another optimization script is used for calculating the required water fluxes. The water flux calculation is done by an approximate model presented elsewhere (Paananen, 2015; Pohjonen *et al.*, 2016). This model calculates the average temperature at the mid-depth of the steel strip. Approximate model was fitted using a more detailed but computationally heavier model (Pyykkönen *et al.*, 2012) so that the effect of strip temperature and water flux was studied with wide variety of parameters in detailed model. These results were used to form tables for mean cooling rate, k, under one cooling unit as a function of temperature and water flux. Later, the mean cooling rate for arbitrary temperature or water flux is calculated by using these tables and linear interpolation.

$$T_{i+1} = T_i + \frac{\Delta z \, l}{t_{strip} v_{strip}} k(T, \theta) \quad (1)$$

where T is temperature, Δz is thickness, l is the length between two modelling points, v is velocity, k is average cooling rate and θ is the water flux. Subscript i denotes a modelling point.

For the induction hardening cases, the optimization time interval was longer, hence it was more suitable to find the initial cooling path by testing few different paths that already produce some amount of the desired phase fractions, so that the optimization script then is capable of finding how the changes of the path affect the resulting phase fractions. While the optimization for the steel strip was fully automated, the semi-automatic optimization was chosen for the induction hardening cases.

3 Results and discussion

3.1 Water cooling of hot rolled steel strip

In the case of steel strip cooling, the chemical composition and rolling procedure limit the possible phase fractions that can be obtained during water cooling. However, in some cases, even small changes to the final phase fractions may lead to desired change in the final properties of steel. Two different cases are presented where the fractions of bainite and martensite are 20 % bainite - 80% martensite for case i) and 40 % bainite - 60 % martensite for case ii). In these cases, five (time, temperature) points were chosen, at which the temperature is optimized. In reality, there could be even more fitting points, but this would make it more difficult to know what the constraints for different zones would be.

First, optimal cooling path to produce desired amounts of bainite and martensite is found through optimization presented in this work, then the water fluxes that produce this cooling path in the cooling line

are calculated with an approximate model (Paananen, 2015). Finally, the cooling path that can be achieved in the cooling line (continuous line in Figures 4 and 6) was used as an input to phase transformation model, to check that even though there are some small differences in the thermal path of Figs 3 and 4, the desired phase fractions are still obtained.

3.1.1 Water cooling of steel strip case i), 20 % bainite, 80 % martensite

The optimized thermal path and phase fractions for water cooled steel strip test case i) are shown in Fig. 3. The path started from fully austenitic structure. The path was optimized to produce 20 % bainite and 80 % martensite. The optimized normalized water fluxes are shown in Fig. 4.
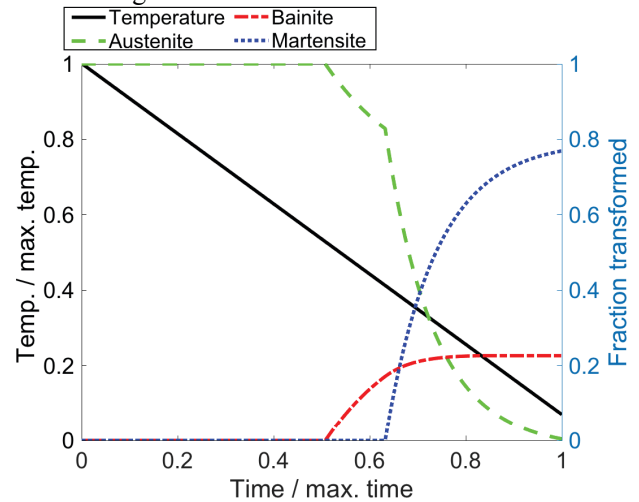
Figure 3. The optimized thermal path for strip cooling process test case i) (20 % bainite, 80 % martensite) and the fraction of phases formed as function of time.
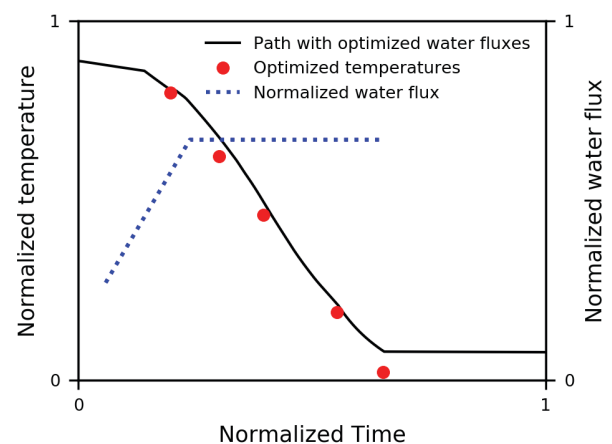
Figure 4. The optimized normalized water fluxes.

Target temperatures refer to temperatures that would be optimal to obtain the desired phase fractions the black line shows the thermal path estimated as result from these water fluxes.

3.1.2 Water cooling of steel strip case ii), 40 % bainite, 60 % martensite

The optimized thermal path and phase fractions for water cooled steel strip test case ii) are shown in Fig. 5. The path started from fully austenitic structure and it was optimized to produce 40% bainite and 60 % martensite. The optimized normalized water fluxes are shown in Fig. 6.
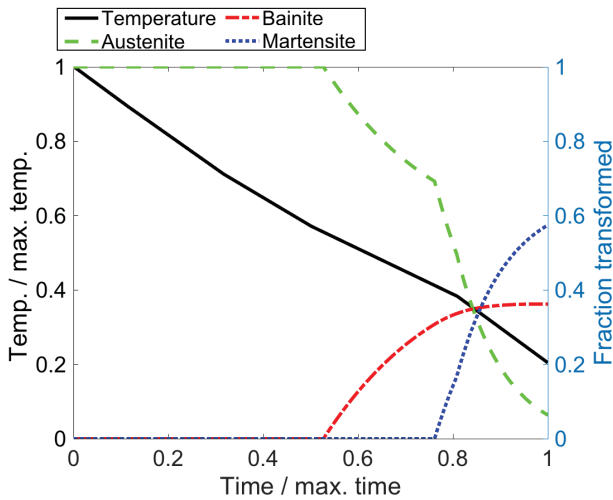
Figure 5. The optimized thermal path for strip cooling process test case ii) (40 % bainite, 60 % martensite) and the fraction of phases formed as function of time.
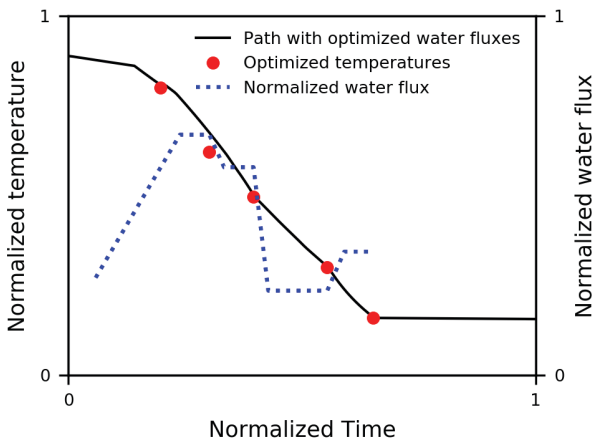
Figure 6. The optimized normalized water fluxes. Target temperatures refer to temperatures that would be optimal to obtain the desired phase fractions the black line shows the thermal path estimated as result from the optimized water fluxes.

3.2 Induction hardening

Typically, in the induction hardening process, in order to avoid a significant grain growth, the time when the material stays in highest temperature (the *peak temperature*) is few seconds. This time is called *dwell time*. In this study, a dwell time of 2.5 s was considered for all the calculations. Moreover, to achieve the finest

possible grain structure, it is desirable to maintain the peak temperature as low as possible, while still achieving full austenitization.

For the optimization of the austenitization, we defined the objective function f_{obj}, which is minimized. The function $f_{obj} = |\chi_\gamma - \chi_{\gamma,des}| + T_p/a$, where $\chi_{\gamma,des} = 1.0$ is the desired austenite fraction (full austenitization). T_p is the peak temperature and a is a penalty parameter which affects how much the peak temperature contributes to the objective function. The parameter a was optimized so that the final value was the minimum that produced more than 99.5 % austenite during the heating and holding in the peak temperature. The physical austenitization parameters were the peak temperature and the heating rate, which was constrained between 50 °C/s and 100° C/s.

The minimum peak temperature satisfying this condition was found to be 795 °C with the minimum heating rate 50 °C/s. The found peak temperature, heating rate, and the chosen 2.5 s dwell time were used in the austenitization schedules for all of the induction hardening cases examined in this study.

3.2.1 Induction hardening case i), fully martensitic structure

The optimized thermal path for the induction hardening test case ii) is shown in Fig. 7. The path was optimized for full austenization at minimum temperature, 2.5 s dwell time, with 50-100 °C/s heating rate. Upon cooling the path was optimized to produce fully martensitic structure. For further experimental and applied studies, as well as serving as a suitable initial guess for other desired phase amounts close to this result, the actual values and cooling rates are presented in Table 1.
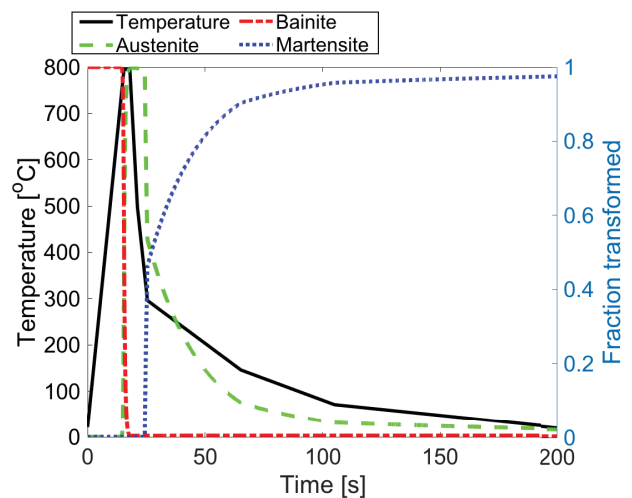
Figure 7. The optimized thermal path for induction hardening process test case i) (fully martensitic structure) and the fraction of phases calculated to form as function of time.

Table 1. The values of the time-temperature path of induction hardening test case i) shown in Fig. 7, as well as the corresponding heating (HR,+) and cooling rates (CR,-).

Time (s)	Temperature [°C]	HR+/CR- [°C/s]
0,0	21,0	50,0
15,5	795,2	0,0
18,0	795,2	-92,2
21,2	497,6	-47,3
25,5	295,9	-3,8
65,3	145,8	-1,9
105,1	70,8	-0,5
200,0	21,0	-

Table 2. The values of the time-temperature path of induction hardening test case ii) shown in Fig. 8, as well as the corresponding heating (HR,+) and cooling rates (CR,-).

Time [s]	Temperature [°C]	HR+/CR- [°C/s]
0,0	21,0	50,0
15,5	795,2	0,0
18,0	795,2	-62,5
20,3	650,3	-2,3
60,9	555,9	-0,2
99,4	548,0	-3,0
149,2	399,0	-7,4
200,0	21,0	-

3.2.2 Induction hardening case ii), equal amount of upper and lower bainite (50% each)

The optimized thermal path for the induction hardening test case ii) is shown in Fig. 8. The path was optimized for full austenization at minimum temperature, 2.5 s dwell time, with 50-100 °C/s heating rate. Upon cooling the path was optimized to produce 50% upper bainite (which forms above 550 °C), 50 % lower bainite (which forms below 550 °C). For further experimental and applied studies, as well as serving as a suitable initial guess for other desired phase amounts close to this result, the actual values and cooling rates are presented in Table 2. Despite the minimization procedure, about 10 % of the austenite transformed to martensite during cooling. A longer total time would be required to produce the desired fully bainitic microstructure.
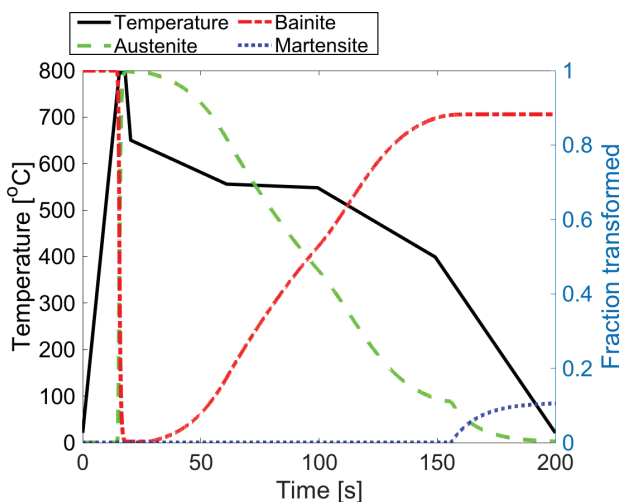
Figure 8. The optimized thermal path for induction hardening process test case ii) (50 % upper and 50 % lower bainite) and the fraction of phases calculated to form as function of time.

3.2.3 Induction hardening case iii), 30% martensite, 40% upper bainite, and 30% lower bainite

The optimized thermal path for the induction hardening test case iii) is shown in Fig. 9. The path was optimized for full austenization at minimum temperature, 2.5 s dwell time, with 50-100 °C/s heating rate. Upon cooling the path was optimized to produce 40% upper bainite (which forms above 550 °C), 30 % lower bainite (which forms below 550 °C) and 30 % martensite. For further experimental and applied studies, as well as serving as a suitable initial guess for other desired phase amounts close to this result, the actual values and cooling rates are presented in Table 3.
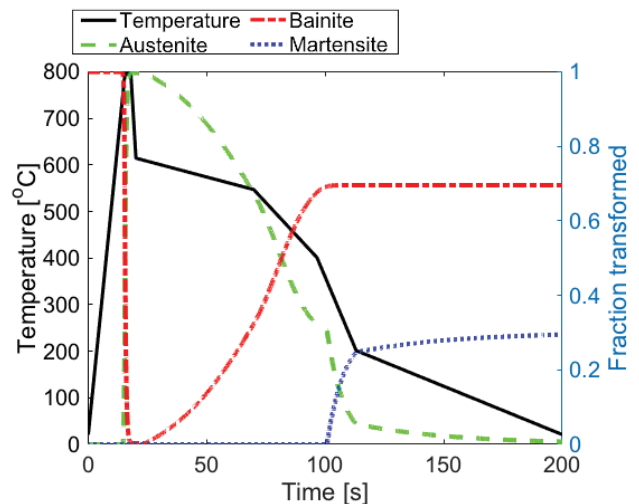
Figure 9. The optimized thermal path for induction hardening process test case iii) (30 % martensite, 40 % upper bainite, 30 % lower bainite) and the fraction of phases formed as function of time.

Table 3. The values of the time-temperature path of induction hardening test case iii) shown in Fig. 9, as well as the corresponding heating (HR,+) and cooling rates (CR,-).

Time [s]	Temperature [°C]	HR+/CR- [°C/s]
0,0	21,0	50,0
15,5	795,2	0,0
18,0	795,2	-84,4
20,1	614,5	-1,4
69,8	547,2	-5,5
96,5	401,1	-12,1
113,1	201,4	-2,1
200,0	21,0	-

4 Conclusions

Using the Matlab programming language (Mathworks, 2020), based on the algorithm described in (Byrd et al., 2000) an optimization script was created for finding the cooling path that produces the desired microstructural constituents (ferrite, upper and lower bainite, martensite) as well as finding suitable heating path for desired austenitization. The script was succesfully applied for several test cases related to water cooling of hot rolled steel strip and induction hardening. For the water cooling case, the script was fully automated by running a series of initial guesses for the thermal path and choosing the best result, while for the induction hardening cases, a user defined initial guess was used. For water cooling we additionally optimized the required cooling water fluxes needed to realize the optimized cooling path based on the earlier work (Paananen, 2015; Pohjonen et al., 2016). Although the results of the presented test cases are not directly applied in the industrial practice as such, they are reasonably close to the actual production to demonstrate the practical applicability of the method.

5 Author contributions

Aarne Pohjonen: implementation of the thermal path optimization script, thermal path optimization simulations, writing of the article.
Vahid Javaheri: original idea of thermal path optimization on induction hardening process, induction hardening process related information, writing, editing.
Joni Paananen: implementation of the water flux optimization tool, water flux optimization, strip cooling related information, writing, editing
Juha Pyykkönen: strip cooling related information, writing, editing.

6 Acknowledgments

References

R. H. Byrd, J. C. Gilbert, and J. Nocedal. A trust region method based on interior point techniques for nonlinear programming. *Mathematical Programming, Series B*, 89:149-185, 2000. doi: 10.1007/PL00011391.

V. Javaheri, S. Kolli, B. Grande, and D. Porter. Insight into the induction hardening behavior of a new 0.40% C microalloyed steel: Effects of initial microstructure and thermal cycles. *Materials Characterization*, 149:165–183, 2019a. doi: 10.1016/j.matchar.2019.01.029.

V. Javaheri, A. Pohjonen, J. I. Asperheim, D. Ivanov, and D. Porter. Physically based modeling, characterization and design of an induction hardening process for a new slurry pipeline steel. *Materials & Design*, 182:108047, 2019b. doi: 10.1016/j.matdes.2019.108047.

M. Jung, S-J. Lee, W-B. Lee, and K.I. Moon. Finite Element Simulation and Optimization of Gas-Quenching Process for Tool Steels. *Journal of Materials Engineering and Performance*, 27(8):4355–4363, 2018. doi: 10.1007/s11665-018-3492-6.

Mathworks. Matlab documentation, 2020. *https://www.mathworks.com/help/matlab/.*

J. Paananen, *Laskennallinen työkalu kuumavalssatun teräsnauhan jäähdyttämisen suunnitteluun*. University of Oulu, BSc thesis. 2015.

A. Pohjonen, M. Somani, J. Pyykkönen, J. Paananen, D. Porter. The onset of the austenite to bainite phase transformation for different cooling paths and steel compositions. *Key Engineering Materials*, 716:368-375, 2016. doi: 10.4028/www.scientific.net/KEM.716.368.

A. Pohjonen, A. Kaijalainen, J. Mourujärvi, and J. Larkiola. Computer simulations of austenite decomposition of hot formed steels during cooling. *Procedia Manufacturing*, 15:1864–1871, 2018a. doi: 10.1016/j.promfg.2018.07.203.

A. Pohjonen, M. Somani, and D. Porter, Modelling of austenite transformation along arbitrary cooling paths. *Computational Materials Science*, 150:244-251, 2018b. doi: 10.1016/j.commatsci.2018.03.052.

J. Pyykkönen, M. Somani, D. Porter, M. Holappa, T. Tarkka. Modelling of Thermal History and Microstructural Evolution on the Run-out Table of a Hot Strip Mill. In *Proceedings from the 6th International Quenching and Control of Distortion Conference*, 9-13 September, 2012 Chicago, IL, USA, pages 817–828, 2012.

R. Tavakoli, Thermal optimization of the continuous casting process using distributed parameter identification approach—controlling the curvature of solid-liquid interface. *The International Journal of Advanced Manufacturing Technology*, 94(1):1101–1118, 2018. doi: 10.1007/s00170-017-0978-6.

Effect of Anisotropic Growth and Grain Boundary Impingement on Bainite Transformation Models

Oskari Seppälä   Aarne Pohjonen   Jari Larkiola

Materials and Mechanical Engineering, University of Oulu, Finland, {oskari.seppala, aarne.pohjonen, jari.larkiola}@oulu.fi

Abstract

In the current study a previously developed cellular automata (CA) model is applied to test how the shape of the growing bainite sheaves as well as the nucleation and growth rates affect the Kolmogorov-Johnson-Mehl-Avrami (KJMA), Austin-Rickett (AR) and the generalized Lee-Kim (LK) model, which encompasses both the KJMA and AR models. The KJMA, AR and LK models are applied to describe the mean field evolution of the phase transformation kinetics. The shape of the growing sheaf affects the impingement phenomena, which then influences the transformation kinetics, even though the overall growth rate for different shape aspect ratios can be the same. Similarly, the different nucleation and growth rates affect the impingement. Systematic simulation studies were performed with different aspect ratios of the growing sheaves as well as different ratios of nucleation versus growth, to see the effect of these factors on the overall transformation kinetics.

Keywords: cellular automata, phase transformations, bainite, steel

1 Introduction

Microstructural evolution is a key factor influencing mechanical properties of steel. Evolution is based on metallurgical phenomena such as austenitization, dislocation processes, recrystallization, grain growth and phase transformations. These phenomena are controlled by process parameters like temperature, including heating and cooling rates, time, deformation and deformation rate. In addition, alloying elements of steel have a strong influence on the metallurgical phenomena. Also, the effect of the preceding process to the subsequent process stage should be included in simulations. In a complete model all these factors need be considered. However, in order to be able to construct an accurate complete model, it is necessary to understand each separate part in detail. The development cycle of new steel grades can be minimized if the final microstructure of high strength steel after TMCP (Thermo-mechanically controlled process) can be predicted. To understand how the nucleation and growth as well as the impingement of the growing bainite sheaves affect the resulting microstructure, we utilize a cellular automata (CA) model to simulate the development of bainite formation (Seppälä et al., 2018).

Although the macroscopic mean field kinetic models, such as the Kolmogorov-Johnson-Mehl-Avrami (KJMA), Austin-Rickett (A-R) or Lee-Kim (LK) models, can be fitted to experimental phase transformation data, they do not provide information on the actual microstructure. The morphology of bainite affects the kinetics because the growing bainite sheaves are obstacles for the growth of other sheaves. To simulate mesoscopic evolution of microstructure, different methods such as phase field (Militzer, 2007), CA (Militzer, 2007) and Monte Carlo (MC) (Kooi, 2006; Militzer, 2007) methods have been used. In the current study an MC-CA model has been applied, which has been developed earlier by authors (Seppälä et al., 2018) for the bainite transformation simulations, because it is capable of simulating the relevant features of the impingement phenomenon.

The CA model is used to study how the bainite sheaf shape affects the resulting morphology and the mean field kinetics described by the KJMA and A-R models. Similarly, the ratio of nucleation to growth rate on the resulting morphology and the mean field kinetics models is investigated.

2 Theory and Calculations

2.1 Mean field model

The mean field theory of bainitic transformation describing the impingement phenomena is examined by fitting the mean field model to the results obtained from the Cellular Automata (CA) model.

Perhaps the most widely used mean field model is the Kolmogorov-Johnson-Mehl-Avrami (KJMA) model, which originally introduced the extended volume concept, i.e. the volume that would be transformed if the transforming regions would not impinge with each other. If V_α^e is used to denote the extended volume and V is the total volume, then we denote corresponding extended fraction by $\chi_e = V_\alpha^e/V$. In the KJMA model it is assumed that the actual transformed fraction χ is related to the extended fraction by the Equation (1)

$$\frac{d\chi}{d\chi_e} = (1 - \chi) \qquad (1)$$

Based on continuous nucleation and growth, the extended volume fraction is calculated as $\chi_e =$

$S \int_{\tau=0}^{\tau=t} \frac{N(\tau)}{V} G(t - \tau)^w \, d\tau$, where S is a shape factor related to the geometry of the growing phase, $\frac{N(\tau)}{V}$ is the nucleation rate per unit volume and $G(t - \tau)$ is the growth rate of the bainite sheaf nucleated at time τ and w describes the dimension of the growing phase. For planar growth $w = 2$. If nucleation and growth rates at the remaining volume stay constant, the extended volume becomes $\chi_e = kt^n$ where $k = \frac{N}{V} G$ and $n = w + 1$. In this case the transformed volume can be written in the widely used closed form, $\chi = 1 - \exp(-kt^n)$. Even if the nucleation rate would decrease, it is still possible to fit the model reasonably well, but with a smaller value of n. (Christian, 1965)

As described in (Starink, 1997), a slightly different transformation model is obtained starting from Equation (2)

$$\frac{d\chi}{d\chi_e} = (1 - \chi)^2 \qquad (2)$$

Again, assuming that the nucleation and growth rates stay constant in the remaining volume, a closed form solution is obtained: $\chi = \left(1 - \frac{1}{1+kt^n}\right)$. This form is the Austin-Rickett (AR) model (Austin and Rickett, 1938).

According to (Starink, 1997), Lee and Kim (Lee and Kim, 1990) proposed a generalized form

$$\frac{d\chi}{d\chi_e} = (1 - \chi)^{1+c} \qquad (3)$$

which includes the parameter c to describe the impingement. For the value $c = 0$ the equation yields the KJMA model, and for $c = 1$ the equation yields the AR model.

In the current study we write the equation (3) in the form $\frac{d\chi}{dt} = (1 - \chi)^{1+c} \frac{d\chi_e}{dt}$, where $\frac{d\chi_e}{dt} = nkt^{n-1}$, which we use to describe the transformation kinetics by applying difference approximation: the transformed fraction $\chi = \sum_i \Delta\chi$ where $\Delta\chi = \frac{d\chi}{dt} \Delta t$ is calculated for each simulation step, and Δt is the simulation time step. The parameters c and n are fitted to transformed fraction calculated with the CA model. By changing the parameters in the CA model, their effect on the impingement parameter c and the Avrami exponent n can be seen.

2.2 CA model

Cellular Automata (CA) is a method where a simulation area is divided into equal-sized squares, cells. These cells each have a user-defined starting value, and these values can be changed by first creating neighborhood rules for the cells to follow and then applying these rules to each cell during simulation steps, which commonly depict the flow of time.

The Cellular Automata (CA) model used here is developed and first presented by authors in (Seppälä et al. 2018). Cell value depicts the phase and instance that it belongs to. In this case, the starting value of each cell is an austenitic instance, grain. During the simulation, new bainitic instances, sheaves, are initiated in austenitic grain boundaries using a probabilistic algorithm. These instances are given a growth vector and a phase instance identifier, and they gradually start to grow in the given direction with a sub-unit -like growth pattern. Figure 1 shows a simplified algorithm schema for bainitic growth. For more details on the algorithms and the equations, see (Seppälä et al., 2018).
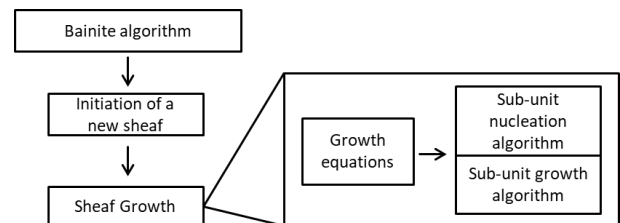
Figure 1. Cellular Automata algorithm schema for bainitic growth.

2.3 Simulations and fitting

The presented CA simulations are not based on experimental results. Instead the starting microstructure is generated with a self-made python script which was initially based on a static recrystallization and grain growth CA model presented in (Sieradzki and Madej, 2013). Other simulation parameters are chosen to simulate a general bainite growth simulation in constant temperature.

The parameters chosen for this study are bainite sheaf growth aspect ratio L_{su} and bainitic nucleation-growth ratio R_b. They are the main factors that can be used to influence the bainite transformation kinetics as well as the morphology of the final microstructure. The values used for fitting are $L_{su} = 4, 6, 8, 10, 12, 14, 16$ and 18 and $R_b = 0.1, 0.5, 1.0, 1.5, 2.0, 2.5$ and 3.0.

The mean field model is fitted directly into the transformation kinetics of each CA simulation case. This way the possible transformation retarding effect of L_{su} and R_b can be linked into the physical parameters of the mean field model. The chosen fitting method is minimization of datapoint difference with equation $S_{comb} = \Sigma|D_{CA} - D_{MF}|$, where S_{comb} is the sum of differences, D_{CA} is a datapoint in the CA simulation and D_{MF} is the mean field datapoint corresponding to the specific CA datapoint. A minimization algorithm finds the smallest possible sum of differences by changing the mean field model parameters. In this study, the truncated Newton (TNC) algorithm was used. It was implemented with Python 2.7, using the optimize.minimize function of module Scipy.

Several fitting limit combinations were tested as shown in table 1. Fitted parameters are n and c. When the parameter name is followed by "fit", the value is fitted with the limits found on the brackets in the table. When the parameter is fixed, the name is followed by

the fixed parameter value. From here on the fitting limit combinations will be referred to with the naming convention presented in the table.

Table 1. Mean field model fitting parameter limits.

Limits	n-3 c-fit	n-fit c-fit	n-2.93 c-fit	n-fit c-0	n-fit c-1
n	3	[2,3]	2.93	[2,3]	[2,3]
c	$[-\infty, \infty]$	$[-\infty, \infty]$	$[-\infty, \infty]$	0	1

3 Results and discussion

3.1 CA model parameters

Before lunging into the mean field parameters, the effect of the chosen CA model parameters R_b and L_{su}. Figure 2a) shows the effect of the parameter controlling bainite sheaf growth aspect ratio (L_{su}) on ferrite transformation kinetics. Increasing the parameter, i.e. emphasizing length growth, slows the later stages of transformation considerably. Figure 2b) shows the effect of the ratio of nucleation to growth rate, controlled by parameter R_b. It seems to have no clear effect on the phase transformation kinetics.

Figure 3 shows effective grain size fraction distributions with a) constant R_b and b) constant L_{su}. Included are example grain structure images illustrating the effect of varying parameters. Increasing R_b seems to steer the distribution towards a larger amount of smaller bainite sheaves. L_{su} does not seem to have a clear effect on the distribution figure. The images show that sheaves emphasize length growth with larger values, which is to be expected.

3.2 Mean field model parameters

The generalized mean field model, described in the theory-section, has two parameters suitable for fitting. First is the Avrami exponent n, and second is the
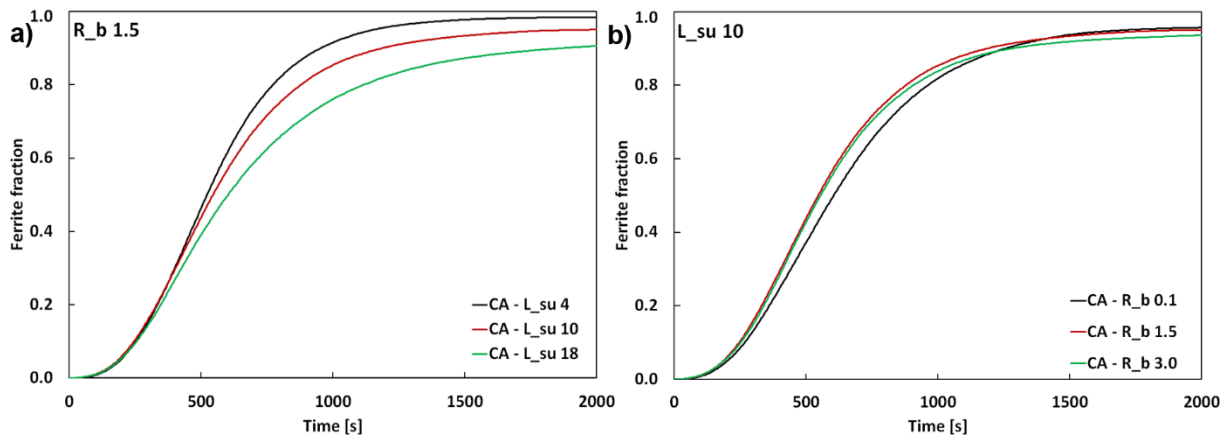
Figure 2. Bainite growth kinetics with CA model using a) constant $R_b = 1.5$ with $L_{su} = 4 \ldots 18$ and b) constant $L_{su} = 10$ with $R_b = 0.1 \ldots 3.0$.
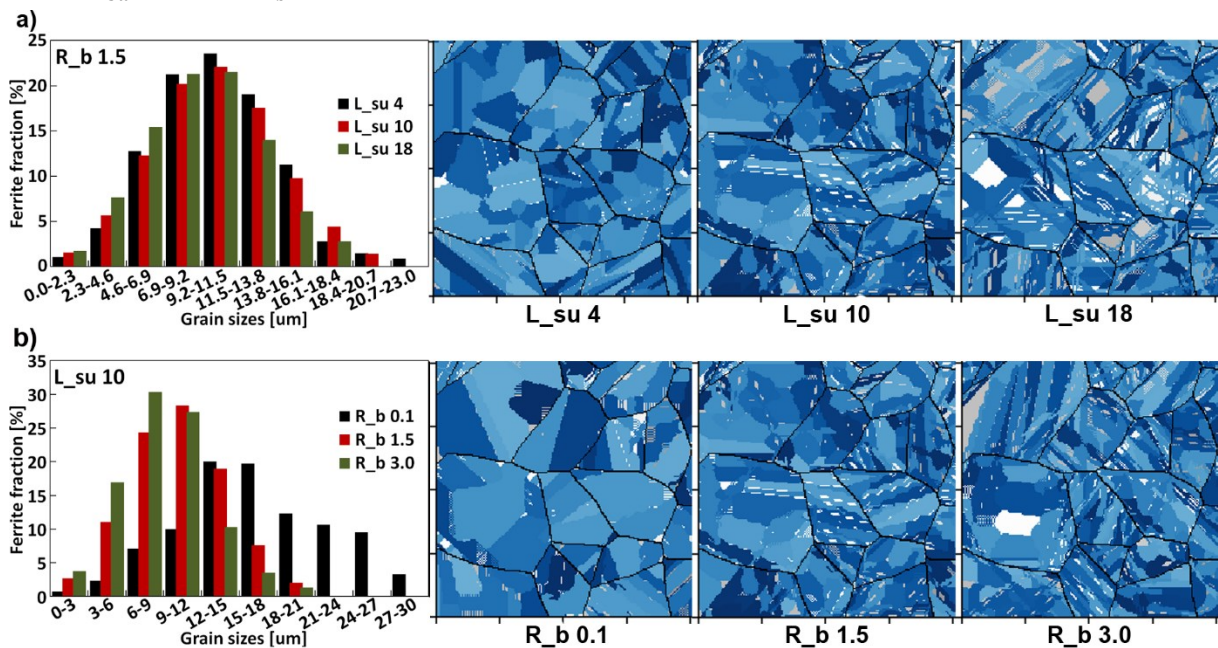
Figure 3. Effective grain size distributions and 100x100 µm sample images for bainite sheaves obtained with the CA model with a) constant $R_b = 1.5$ with $L_{su} = 4 \ldots 18$ and b) constant $L_{su} = 10$ with $R_b = 0.1 \ldots 3.0$.

impingement factor c. Figure 4a) includes a few selected parameter values to illustrate their effect on the transformation kinetics. Figure 4b) shows the accuracy of the chosen fitting limits with an example case where $R_b = 1.5$ and $L_{su} = 10$. The classic KJMA equation has the impingement factor c set to 0. This clearly sets it apart from the other fitting limits, which have only minor differences.
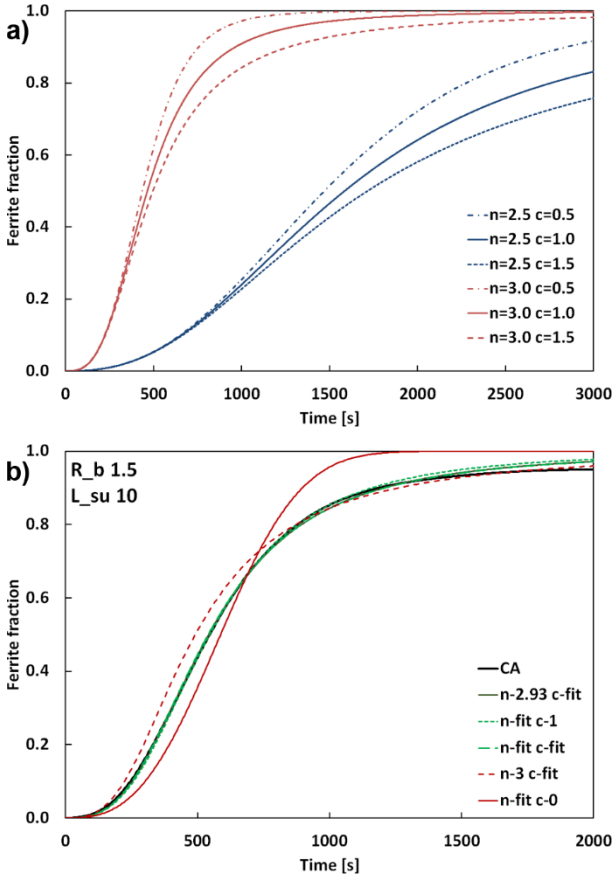
Figure 4. a) Illustrative figure of the effect of fitting parameters b) Different fitting limit examples.

A more comprehensive look on the fitting curves is provided in Figure 5. It shows the CA model phase transformation kinetics with $R_b = 1.5$ and $L_{su} = 6, 10, 14$ and 18. Fitting limits n-fit+c-fit were used. The correspondence between the CA results and the fitted mean field models is quite good and based on this the fitting can be called successful. The early stages of phase transformation correspond perfectly to CA data, but the later stages show minor variations. One potential problem is that the mean field model exaggerates phase transformation at the very end of the curve. The difference would be somewhat larger if the curves were to be extrapolated further. The fitting could be improved by emphasizing the end stages in the fitting procedure.
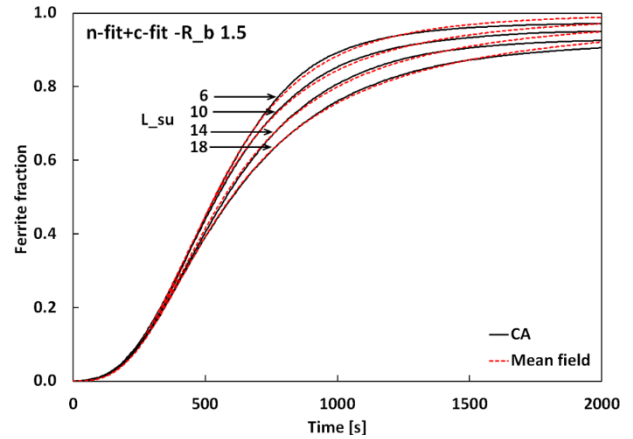
Figure 5. Mean field model fitting to CA with the n-fit+c-fit limits and $R_b = 1.5$.

The fitting parameters obtained with varying fitting limits are compared against each other by L_{su} and R_b. Figure 6 includes the fitted values for a) Avrami exponent n and b) impingement factor c. Only the limits where the corresponding parameter is fitted are included in the figures. The values are grouped by R_b as function of L_{su}. For both parameters it can be seen that R_b shows no clear effect, unlike L_{su} which shows quite a nice linear dependence for both parameters. The only notable exception is n-fit+c-fit, where n seems to be almost a constant value. For this reason, another limit case was chosen with $n = 2.93$. The chosen value is the average of n-values from the n-fit+c-fit limits, excluding $R_b = 0.1$ which is considered an unreliable parameter value due to the abnormal size of the sheaves produced, considering the morphology of bainite.
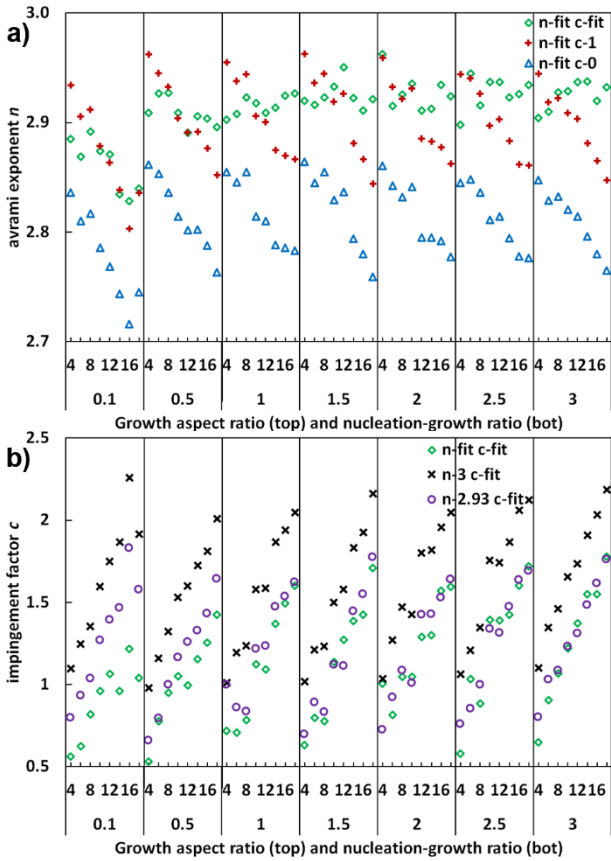
Figure 6. Fitting parameters a) n and b) c, sorted by growth aspect ratio L_{su} and nucleation-growth ratio R_b.

Figure 7 shows the final sums of differences for all fitting limits with a similar figure structure as in Figure 6. The sum of differences can be used to compare the fitting limits with each other and estimate how well the fitting succeeded. Some variations are to be expected, so all the fittings can be considered successful apart from n-fit+c-0, which is the classical KJMA equation. Not taking impingement into account shows a clear deterioration of fitting accuracy with increasing L_{su}-values.
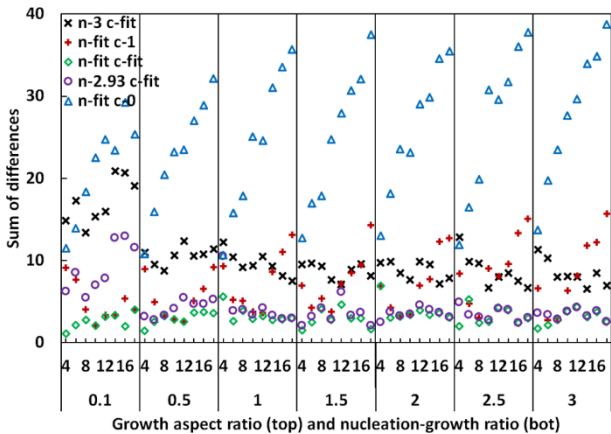
Figure 7. Comparison of the fitting accuracy of different configurations using sum of differences.

3.3 Parameter quantification

The results in Figure 6 are very clearly linearly dependent on L_{su}. To quantify this dependence, the results were fitted linearly with the equation $y = k_y L_{su} + y_0$, where y is c or n, k_y is the angular coefficient and y_0 is the intersection of the linear fit with the y-axis in Figure 8. Figure 8 shows fitting curves and table 2 shows linear fitting results for all R_b-values for the impingement factor c with fitting limits n-2.93+c-fit.
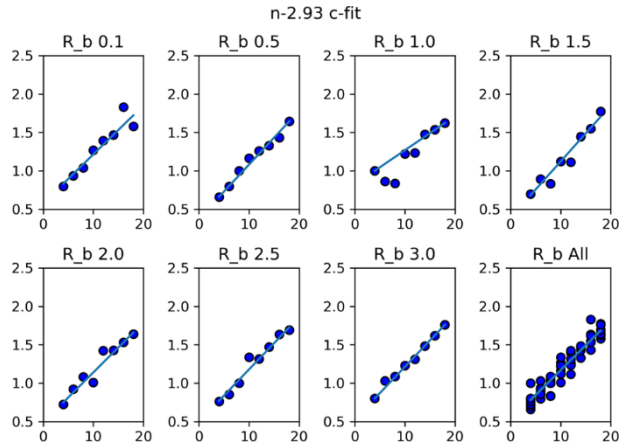
Figure 8. Linear fitting for impingement factor c as function of L_{su} with varying R_b for the fitting limits n-2.93+c-fit.

Table 2. Linear fitting parameters for the impingement factor c with limits n-2.93 c-fit.

R_b	0.1	0.5	1	1.5
y_0	0.576	0.379	0.822	0.390
k_y	0.064	0.070	0.045	0.073
R_b	2	2.5	3	All
y_0	0.487	0.496	0.525	0.522
k_y	0.065	0.068	0.069	0.065

The effect of R_b on the transformation kinetics seems to be mostly random based on the results so far. To see any quantifiable difference between R_b-values, the fitting results were combined into averaged values of y_0 and k_y. These results and their sums of differences are presented for both c and n in Tables 3 and 4. Standard deviations are also included to estimate the accuracy of the averaging.

Table 3. Linear fitting parameters for the impingement factor c for all R_b.

Limits	y_0 avg (stdev)	k_y avg (stdev)	S_{comb} avg (stdev)
n-3 c-fit	0.787 (0.083)	0.074 (0.005)	0.390 (0.175)
n-2.93 c-fit	0.525 (0.149)	0.065 (0.009)	0.448 (0.199)
n-fit c-fit	0.409 (0.190)	0.067 (0.017)	0.590 (0.163)

Table 4. Linear fitting parameters for the Avrami exponent n for all R_b.

Limits	y_0 avg (stdev)	k_y avg (stdev)	S_{comb} avg (stdev)
n-fit c-0	2.877 (0.012)	-0.00621 (0.00101)	0.055 (0.016)
n-fit c-1	2.979 (0.013)	-0.00720 (0.00085)	0.055 (0.017)
n-fit c-fit	2.911 (0.027)	-0.00002 (0.00248)	0.074 (0.027)

All the fitted parameters in tables 3 and 4 show only small differences in the standard deviations, with the notable exception of n-fit+c-fit, where the standard deviation of y_0 and k_y is larger. Also, the sums of differences (S_{comb}) show relatively large standard deviations. These remarks combined indicate towards the idea that R_b may have only a random effect on the time evolution of the overall bainite fraction. It is not enough to confirm this idea though, so the subject should be investigated further. However, this does not adversely affect the general applicability of the L-K model in describing the impingement phenomenon.

4 Conclusions

Based on the results presented here, some remarks can be made about the effects of the CA parameters chosen for investigation. Firstly, the nucleation to growth ratio, R_b, has a significant effect on the resulting effective grain size distribution but not on the phase transformation kinetics. These findings agree with the used nucleation and growth theory, and R_b will be a useful parameter when fitting the model using experimental microstructures.

Secondly, L_{su} has no discernible effect on the effective grain size distributions, shown in Figure 3a). This is to be expected, since calculating effective grain size eliminates mathematically any anisotropy from bainite sheaves. More work needs to be done to find a method that properly quantifies the effect of growth anisotropy on simulated bainite sheaf morphology.

L_{su} has a significant effect on the phase transformation kinetics. Higher values decelerate phase transformation. This is most likely caused by growing sheaves colliding into austenite grain boundaries and other bainite sheaves, which occurs faster for sheaves with more longitudinal growth. This is the impingement effect.

The generalized mean field model has been fitted into the CA model with c and n. The fitting results showed good agreement with the CA simulation results, and a clear correspondence between L_{su} and c was found. In the currently applied theory, c is the factor controlling impingement. The results of this study support the idea that L_{su} has a major impact on the impingement during the simulation. The effect of impingement was quantified for the mean field model with a linear fitting, which also showed quite good correspondence.

Within the used fitting limits some cases showed correspondence between L_{su} and n as well, but when both parameters were fitted simultaneously, n showed only minor variations. Based on the used theory, n is expected to vary based on the diminishing amount of nucleation sites in the simulation area and the results presented here support that sentiment.

In future studies these results can be used to further investigate impingement with experimental results. The CA model will also be developed further. Multiple CA simulations with identical parameters will be run to obtain statistically more reliable results and further study the results shown in this study.

References

J. B. Austin and R. L. Rickett. Kinetics of the Decomposition of Austenite at Constant Temperature. *American Institute of Mining and Metallurgical Engineers* V(Technical publication n.964): 1938.

J. W. Christian. *The Theory of Transformations in Metals and Alloys*. Pergamon Press, 1965

J. B. Kooi. Extension of the Johnson-Mehl-Avrami-Kolmogorov Theory Incorporating Anisotropic Growth Studied by Monte Carlo Simulations. *Phys. Rev. B* 73(5): 2006. doi:10.1103/PhysRevB.73.054103

E. S. Lee and Y. G. Kim. A Transformation Kinetic Model and Its Application to CuZnAl Shape Memory Alloys-I. Isothermal Conditions. *Acta Metallurgica Et Materialia* 38(9): 1669–76, 1990. doi:10.1016/0956-7151(90)90009-6

M. Militzer. Computer Simulation of Microstructure Evolution in Low Carbon Sheet Steels. *ISIJ international* 47(1): 1–15, 2007. doi: 10.2355/isijinternational.47.1

O. Seppälä, A. Pohjonen, A. Kaijalainen, J. Larkiola and D. Porter. Simulation of Bainite and Martensite Formation Using a Novel Cellular Automata Method. *Procedia Manufacturing* 15: 1856–63, 2018. doi:10.1016/j.promfg.2018.07.204

L. Sieradzki and L. Madej. A Perceptive Comparison of the Cellular Automata and Monte Carlo Techniques in Application to Static Recrystallization Modeling in Polycrystalline Materials. *Computational Materials Science* 67: 156–73, 2013. doi:/10.1016/j.commatsci.2012.08.047.

M. J. Starink. Kinetic Equations for Diffusion-Controlled Precipitation Reactions. *Journal of Materials Science* 32(15): 4061–70, 1997 doi:10.1023/A:1018649823542.

Modelling of the Lime Kiln at SSAB, Raahe

Tomas Lundberg[1]   Joel Orre[1]   Timo Paananen[2]   Olli Mattila[2]

[1]Swerim AB, Luleå, Sweden, `{joel.orre,tomas.lundberg}@swerim.se`
[2]SSAB, Raahe, Finland, `{timo.paananen,olli.mattila}@ssab.com`

Abstract

An OpenModelica model of the SSAB Raahe lime kilns has been developed in order to help in formulation of operation strategies and to choose important parameters to measure and monitor. The model is a dynamic simulation model describing the calcination process of limestone in the lime kiln.

Keywords: lime kiln, lime calcination, process modelling, dynamic simulation, dynamic model

1 Introduction

The steel industry is a major consumer of lime products. To produce 1000 kg of crude steel (using the integrated steelmaking route based on Blast Furnace (BF) and Basic Oxygen Furnace (BOF)), about 270 kg of limestone is needed (World Steel Association 2019). The estimated annual world production of lime is around 350 million tons, of which the steel industry uses between 140 and 160 million tons globally (Manocha and Ponchon 2018).

Lime has several usages in the steel making process. In the BF, where iron ore is smelted into pig iron, limestone is used for the formation of slag as it accumulates the waste materials that are produced in the process. During basic oxygen steelmaking, where hot metal is converted into steel, quicklime is used in the BOF process both for phosphorus and sulphur removal and for slag formation.

The lime quality has an important impact on the steel quality, its metallurgical properties, and the total cost of production. If the lime is over-calcinated it starts sintering which leads to a reduction in the surface area and hence loss of reactivity. Similarly, if the lime is under-calcinated, it will still contain carbon dioxide in its core (see Section 2 for further details) which also lowers the reactivity and may also cause further problems during the BOF process. It is therefore important to have control over the lime quality.

An OpenModelica model of the SSAB Raahe lime kilns was developed in this study. The purpose of the model is to help in formulation of operations strategies and to aid in selecting important parameters to measure and monitor. Questions that the model can help answering is e.g. how the lime kilns should be operated during production stops and how to adjust the lime quality. The model is a dynamic simulation model describing the calcination process of limestone in the lime kiln.

In the literature, simulations of lime kilns have mostly been done with one-dimensional approaches (see e.g. (Bes 2006; Do 2012; Herz et al. 2015) and the references therein). More advanced three-dimensional Discrete Element Method-CFD (DEM-CFD) simulation has also been presented (Bluhm-Drenhaus et al. 2010; Krause et al. 2017). However, these simulations are quite computational extensive; the simulation performed by (Krause et al. 2017) required about three weeks computational time on 2016 mainstream hardware. Our simulation is less detailed but doesn't require more than around 100 minutes computational time for about 24 hours simulated time, or about 4 minutes per hour simulated time (on an ordinary office laptop). This is short enough for enabling the model to be used as a practical tool for selecting the best operation strategies.

The paper is organised as follows. In Section 2 we give a short introduction to the calcination process, while Section 3 give a general overview of lime kilns. In Section 4 we describe the OpenModelica model, and Section 5 discusses the calibration of the model against production data. Finally, Section 6 contains a discussion of the results from modelling and we end the paper in Section 7 with the conclusions.

2 Lime Production

Lime is manufactured by burning limestone in kilns. The limestone is crushed into particles with a size ranging from a few centimetres up to around 10 centimetres. In the calcination process calcium carbonate is converted to calcium oxide and carbon dioxide. This is an endothermic process with the following reaction formula

$$CaCO_3 + heat \longrightarrow CO_2 + CaO. \tag{1}$$

The required heat is supplied by burning fuel within the kiln.

On the particle scale, the actual calcination reaction is advancing from the surface of the particle to the centre, a so-called "shrinking core" process. Two steps are determining the rate: the heat transfer from the surrounding gas to the reaction front, and the transfer of the released CO_2 through the particle to the surface where it is released to the surrounding gas.

To model the calcination process we have taken into account details of heat and mass transfer. We mainly have followed the approach by (Do 2012), to which we refer for further details.

3 Lime Shaft Kilns

There are two basic types of lime shaft kilns in use today. A normal shaft kiln has a single vertical shaft where limestone is charged at the top and lime is discharged at the bottom. The lime and limestone move slowly downwards in the shaft by gravity. The calcination occurs in the middle of the kiln where fuel is combusted.

A Parallel Flow Regenerative (PFR) kiln is a more modern type of kiln which has two interconnected shafts (see Figure 1, left). Each shaft operates in two different modes, burning and non-burning, and while one shaft operates in the burning mode (calcinating the limestone), the other operates in the non-burning mode (preheating the limestone). Each shaft can also be divided into three different zones, the pre-heating zone, the combustion zone, and the cooling zone. A cross-over channel connects the two shafts at the bottom of the combustion zone.

In the burning mode, fuel is supplied through lances at the top of the combustion zone. Combustion air is supplied under pressure at the top of the kiln (at the top of the pre-heating zone). The combustion gases pass through the crossover channel from the burning shaft into the non-burning shaft, where they flow upwards and exit at the top of the non-burning shaft. The flow of gases therefore changes direction when the shaft shifts from burning to non-burning. During the passage, the off-gases transfer heat to the limestone in the non-burning shaft. Cooling air is entering the shaft at the bottom of the cooling zone. The flow of gases changes direction at regular intervals of about 15 minutes, which is called a half cycle. Another flow reversal completes the cycle.

Limestone is supplied at the top of the shaft. Lime and limestone travels slowly down through the shaft by gravity. During the non-burning mode, the stone in the pre-heating zone is heated by the combustion gases, while during the burning mode the stone transfers heat to the combustion air. The stone in the pre-heating zone thus acts as a regenerative heat exchanger.

To the right in Figure 1 is the temperatures profile for one of the shafts. The profile should be thought of as following a mass particle from the top of the shaft to the bottom. The vertical axis is therefore a combination of time and location in the shaft. The variation in temperature due to switching between burning and non-burning mode can be seen clearly.

3.1 The Lime Kilns at SSAB Raahe

There are two PFR shaft kilns in Raahe with production capacity of 280–330 ton/day and 300–420 ton/day, respectively. They are operated with different types of limestones sizes, between 40 and 70 mm for lime kiln 1 and between 60 and 90 mm for lime kiln 2. Due to the size distribution of the limestones a good burned lime product still contains about 2.5% $CaCO_3$, which is found in the core of the larger pieces. As discussed in Section 1 both over-calcination and under-calcination is bad. However,

because most of the limestone is used in BOF the worst case is when the quality varies since this gives bad energy predictions.

One difference between the lime kilns at SSAB Raahe and the general description in Section 3 is that the combustion air is not supplied at the top through pre-heating zone; instead it is mixed with Coke Oven Gas (COG) and injected through the lances.

4 Lime Kiln Modelica Model

The lime kiln Modelica model is built up in a modular fashion. At the top level there is a total model for a single shaft including pre-heating, combustion and cooling. Each of these three zones are in turn modelled in separate sub-models. The sub-models will be described further below.

The models use mass and energy balances together with reaction kinetics according to calcination and a simple flame model. SSAB has done tests on the used limestones at one temperature which is used as a base for formulating a shrinking core model of limestone calcination.

4.1 Total Model

The total model is shown in Figure 2. It consists of three sub-models which corresponds to the three zones in a PFR lime kiln. The three sub-models, the pre-heating model, the combustion model, and the cooling model, will be described further below.

In the model we only consider one of the two parallel shafts explicitly. The two distinct modes of operations for a PFR lime kiln, burning and non-burning, are modelled using off-line tables which specify the times for burning and non-burning. The cross-over channel between the two shafts in a PFR kiln is modelled by adding a gas sink and a gas source between the combustion zone and the cooling zone. The sink is used during the burning mode where in a two-shaft PFR the combustion gases would have been used to heat the stone in the non-burning shaft. Similarly, the gas source is used during the non-burning mode to mimic the transfer of combustion gases to the non-burning shaft. The gas flow parameters for the sink and the source are governed similarly as the burning times, i.e. using off-line tables. The data in the tables is taken from process data.

Using a single shaft in the model in this way greatly reduces the complexity of the model, but at the expense of some flexibility in the modelling since it requires process data.

4.2 Combustion Model

We start by describing the combustion model (see Figure 3). The same model (with small modifications) is also used in the other two models, the pre-heating model and the cooling zone model.

The combustion model is used for modelling the middle part of the kiln, where the limestone is decomposed in lime and CO_2. Since the decomposition of limestone
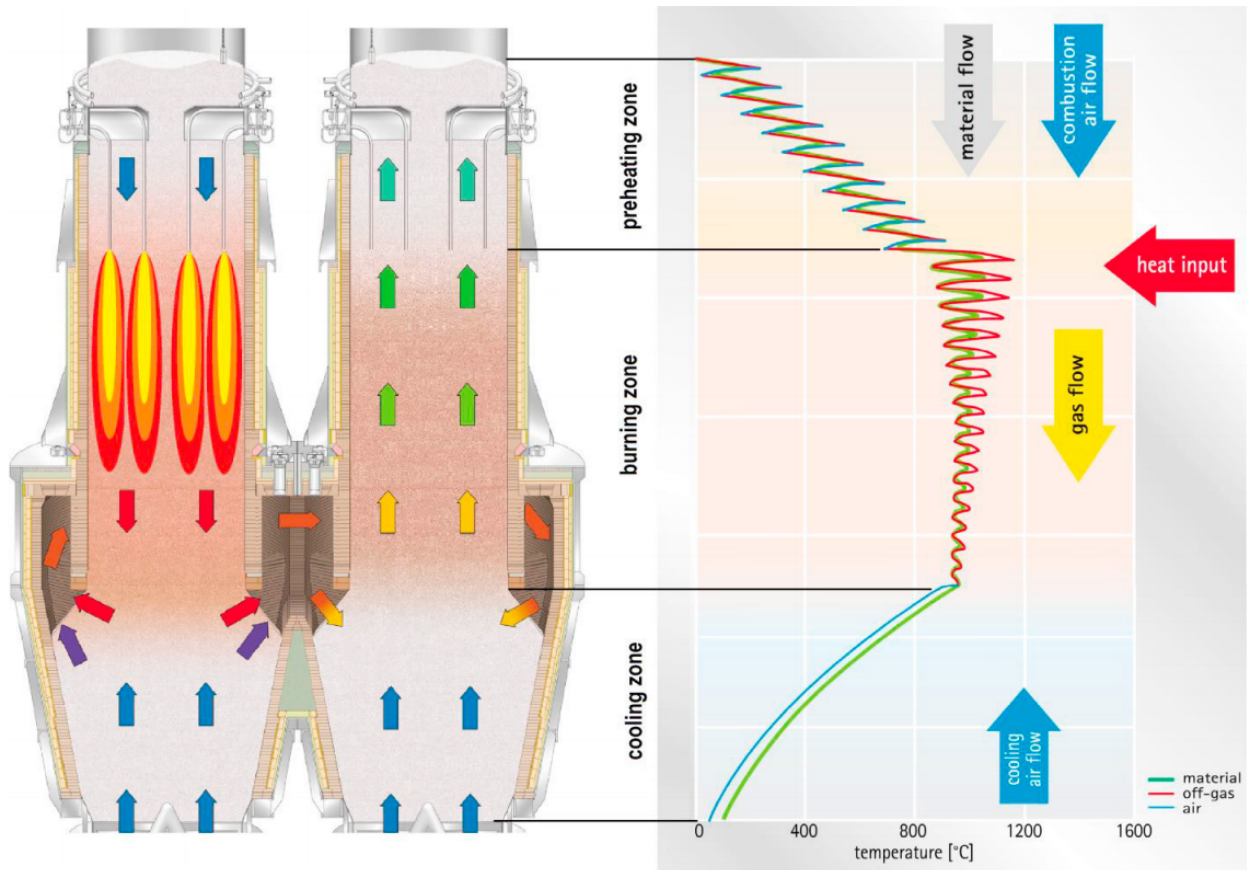
Figure 1. Operating principle and temperature profile of a PFR lime kiln. Taken from (Piringer 2017).

is an endothermic process it needs an energy supply. The heat needed for the calcination is generated by burning fuel (COG) in the combustion zone, and the fuel (together with the combustion air) is introduced in the combustion zone through lances at the top of the zone.

The combustion model is conceptually quite simple. As input it takes the gas and lime/limestone mass flows from the pre-heating model. The output is the resulting gas and lime/limestone mass flows which is used as input to the cooling model. Notice that for the gas flow, "input" and "output" is here used conceptually and should not be interpreted literally. The direction of the actual gas flow varies depending on the mode of the shaft: in the burning mode the direction of the gas flow is from the pre-heating model to the combustion model ("downwards" in this conceptual view), while in the non-burning mode the direction is from the combustion model to the pre-heating model ("upwards").

The core of the combustion model is a limestone model and a packed bed model.

The limestone model is responsible for the actual modelling of the conversion from limestone to lime. This model is based on the shrinking core reaction rate calculations (see e.g. (Do 2012)). The decomposition of limestone into lime releases CO_2 which is fed back to the packed bed model and included in the gas flow.

The packed bed model is responsible for the modelling of the gas flow through the limestone. Additionally, it calculates the heat transfer between the gas and the limestone. The total amount of available heat is calculated from the COG flow based on the heat of combustion. In order to model the physical extent of the flame, the combustion is divided into a number of sections, basically by just stacking identical sections of the combustion model on top of each other. We have used a model with 20 sections.

The actual size of the flame is not known with certainty, and therefore the heat produced by the flame is divided between the sections by specifying the fraction of the total heat produced that is available in a given section. In the simulations we have performed, the flame extends through the 13 topmost sections, and where the highest amount of heat is delivered in the top-most sections and decreasing towards the bottom.

There are some unknown factors which need to be taken into account in the combustion model, such as the heat losses to the surroundings. Another factor is the uncertainty in the measured value of the heating value for the COG. We have taken these factors into account in an empirical way by lowering the measured heating value slightly with approximately 12 %, and thus decreasing the amount of heat delivered as calculated from the COG flow
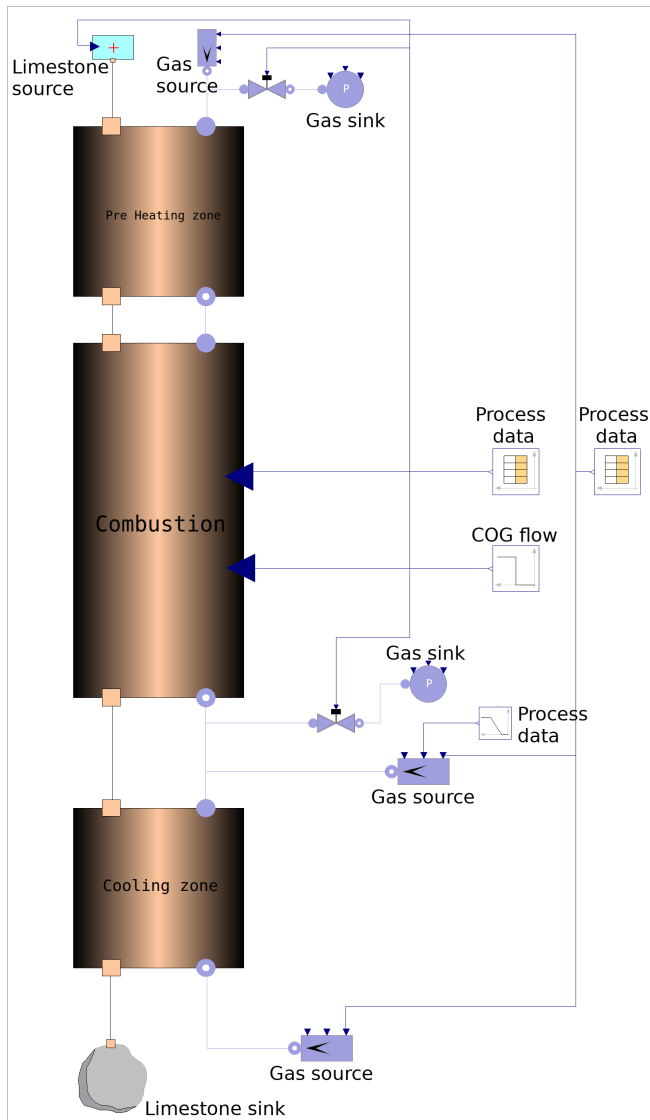
Figure 2. Total lime kiln model.

in the model. The value was determined from fitting the model to process data.

A pressure drop is also inserted in the gas flow to model the gas flow resistance.

4.3 Pre-heating Model

The pre-heating model is in principle identical to the combustion model. Basically, the only difference is that there is no combustion and consequently the COG flow is not used as input. Similar to the combustion model, the pre-heating model is also divided into sections and as in the combustion model we have used 20 sections also in the pre-heating model.

The input to the pre-heating model is the mass flow (i.e. the limestone inserted at the top of the kiln) and gas flow (both derived from tables as described for the total model in Section 4.1).
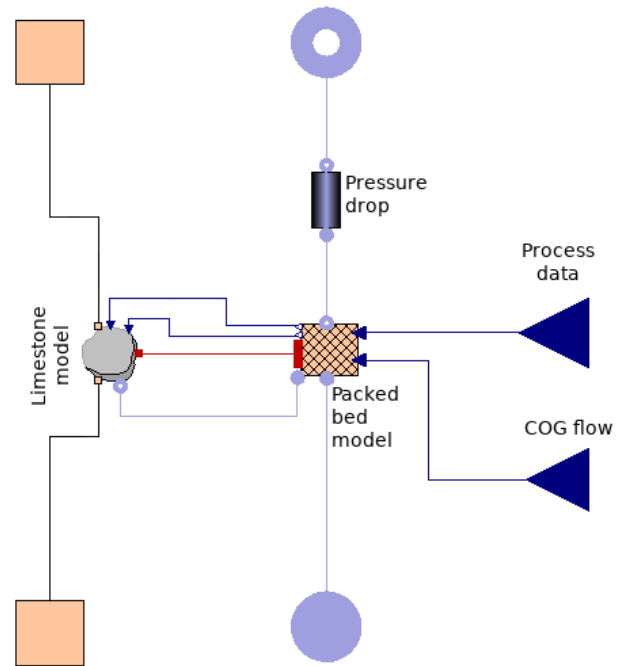
Figure 3. The combustion model. The gas flow is shown in blue and the limestone flow in beige.

4.4 Cooling Model

The model for the cooling zone is identical to the preheating model. The input to the cooling model is the mass flow and gas flow from the combustion model. The finished lime is received by a "lime sink".

5 Model Calibration

The shrinking core model describing limestone calcination is calibrated with data from laboratory tests studying the calcination. The laboratory test were done through preheating a limestone cylinder with a diameter of 20.8 mm, a height of 20.5 mm, and an initial mass of 18.5 g to 800 °C. The limestone was then put into a furnace with constant power applied and with a temperature of 950 °C. In the furnace, the mass of the limestone and the temperature close to the limestone were measured. This laboratory test was then modelled with limestone calcination based on shrinking core model found in (Do 2012), with a slight difference on how the heat conduction is modelled. In our model the heat conduction is modelled outside the shrinking core model for more precise heat transfer. The following constants were calibrated:

- heat conduction constant for limestone: $\lambda = 0.6\,\mathrm{W/(m \cdot K)}$,

- diffusion constant for CO_2: $d_p = 2 \times 10^{-4}\,\mathrm{m^2/s}$,

- reaction rate constant: $k = 8 \times 10^{-3}\,\mathrm{m/s}$.

These constants give similar rate of calcination in the model compared to the laboratory test as seen in Figure 4.
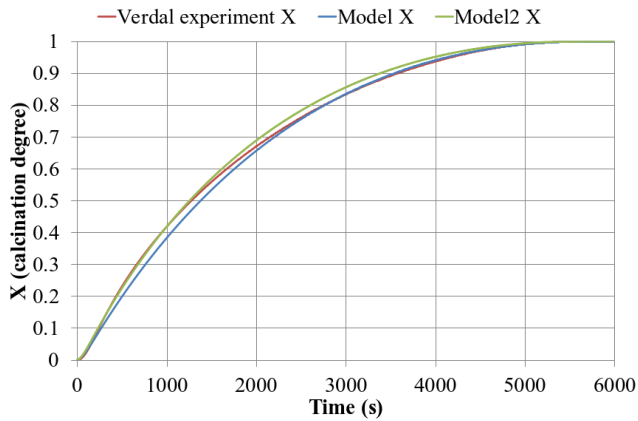
Figure 4. Calcination degree of calibrated model (Model X), updated calibrated model (Model2 X) and laboratory test (Verdal experiment X).

Figure 5. The steady-state conversion profile of limestone (blue) together with the maximum and minimum gas temperature (orange).

The heat conduction constant λ was later changed from 0.6 to 0.55 in order to achieve better calibration with lime kiln process data. This updated λ in Model2 has good correlation to the laboratory test in the first 1500 s (see Figure 4). To achieve even better calibration of calcination constants it would have been preferable to have laboratory tests on at least two different furnace temperatures for the same limestone, but it has not been done in this case.

Concerning the gas-solid heat transfer coefficient we have followed (Do 2012) and references cited therein. The heat transfer coefficient α can be calculated from the Nusselt number in the packed bed (Do 2012). In order to make the energy balance easier, they introduced a modified heat transfer coefficient which assumes a homogeneous average temperature instead of having a temperature distribution in the radial direction in the particle due to that the heating-up and cooling-down of the solid particles is a transient process. Additionally, we added a factor to the α constant for calibrating the heat transfer. Our modified heat transfer coefficient α_{mod} is then given by

$$\alpha_{\mathrm{mod}} = k_\alpha \left(\frac{1}{\alpha} + \frac{d/2}{\kappa\lambda} \right)^{-1}, \quad (2)$$

where k_α is the calibration constant, λ is the thermal conductivity of the solid particle, d the particle diameter, and κ is a transient factor which equals 5 for a sphere. The value for our α_{mod} in the calibrated OpenModelica model ranged from 27 W m^{-2} K to 29 W m^{-2} K, which is reasonably close to heat transfer coefficients found in (Senegačnik et al. 2007). The modified heat transfer coefficient enters in the energy balance equations, e.g.

$$\dot{M}_s c_{ps} \frac{dT_s(z)}{dz} = \alpha_{\mathrm{mod}} A\, O\, (1-\psi)\,(T_s(z) - T_g(z)), \quad (3)$$

where $\dot{M}_s$ is the mass flow of lime, c_{ps} its specific heat capacity, A the cross-sectional area of the furnace, O the specific surface of the lime in m^2/m^3, ψ the void fraction, $T_{s/g}$ the temperature of the lime and the gas, respec-

tively, and z the position along the height of the kiln. In our model we have used a void fraction of 0.38.
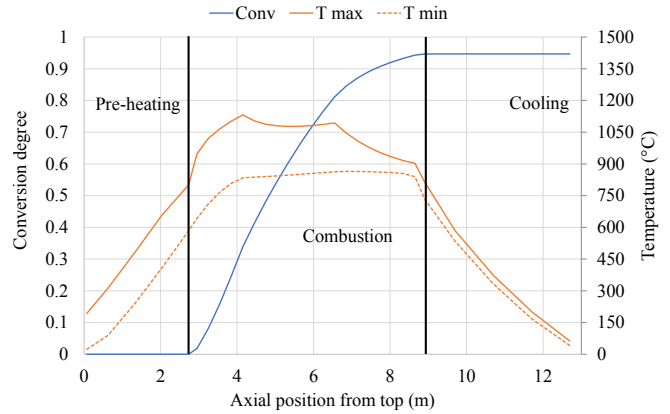
6 Simulation Results and Discussion

We have applied our model to one of the lime kilns in Raahe. From the model it is possible to obtain e.g. the gas temperature and conversion profile of the limestone. Figure 5 shows the steady-state conversion profile, averaged over one complete cycle, together with the maximum and minimum temperature for the gas during one cycle.

As a more advanced example of application for our model, we have used it to study how to best continue operation of the kiln after a stop. We used process data for 24 hours as input, which included a production stop lasting a couple of hours.

The lime kiln operating conditions before and after the stop is similar, i.e. same cycle time and limestone charge.

Figure 6 shows temperature measurements in two locations in the cross-over channel. The temperature decrease in the channel during the stop is around 50 °C to 60 °C. Based on this, the temperature decrease in the modelling is assumed to be 50 °C in large part of the furnace (where the temperature of the lime is over 500 °C).

We have studied three different approaches to start the kiln after a stop. Description of these cases, together with the "Steady state" nominal case, is found in Table 1. In the three cases concerning how to restart the kiln, lime/limestone temperature is assumed to have cooled with 50 °C in the warm part of the lime kiln. No calcination is assumed to take place during the stop, which is a good assumption if there is no gas flow in the lime kiln during the stop. The reason for this is that the partial pressure of CO$_2$ prevents further calcination already when only a small amount of calcination has taken place. On the contrary, with only small flow of air the calcination continues significantly.

The heat loss during this particular stop corresponds to approximately the same amount of heat that is added to the lime kiln during a half cycle, or, equivalently, about 50 % of the COG used in one cycle. The heat loss can therefore

Table 1. Simulation cases.

Case name	Description
Steady state	Based on average values from process data before the stop.
Normal	Simulation of process data conditions after stop. Operation continues in same way as before stop with the change that limestone/lime temperature starts 50 °C lower (in large part of lime kiln).
COG	Limestone/lime temperature starts 50 °C lower than before stop. The first cycle is operated with higher COG flow (150 % of normal operation conditions)
Half cycle	Limestone/lime temperature starts 50 °C lower than before stop. The discharging of lime starts after a half cycle instead of directly after stop. Combustion is done in both shafts with combustion time halved.
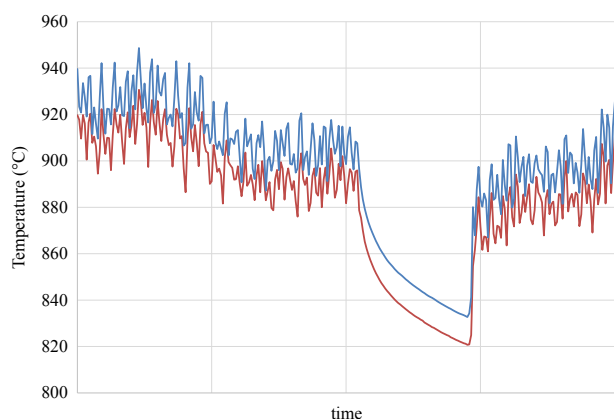
Figure 6. Temperature measurements in the cross-over channel for two different locations. The effect on the temperature by the production stop can clearly be seen.

Figure 7. The percentage of CO_2 in the discharged lime.

be compensated either by delaying the lime discharge for one half cycle (the "Half cycle" case) or by increasing the flow of COG with 50 % for one cycle (the "COG" case).

Figure 7 shows the percentage of CO_2 in the discharged lime while Figures 8 and 9 show the rate of calcination and the average gas temperatures for the four cases, where the values have been averaged over the three first cycles after the stop. Comparison with the steady state conditions show that the percentage of CO_2 in lime for the "Normal" case is increased at 5 h to 22 h after start with almost 1.5 % at its peak.

The "COG" case gives slightly lower %CO_2 in lime at 6 h to 11 h and slightly higher %CO_2 (0.45 % at peak) at 12 h to 20 h after operation has continued after stop. This is due to a lower rate of calcination in the first part of the combustion zone (at 2.8 m to 4 m in Figure 8) and a lower temperature in the pre-heating zone and first part of the combustion zone (at 5 m to 7 m in Figure 9) in the first cycles after stop.

In the "Half cycle" case there is a slight increase (0.35 %) in the %CO_2 at 5 h to 9 h and a slight decrease (−0.3 %) in %CO_2 at 10 h to 16 h. The reason for this is a warmer and more calcined upper part of the lime kiln and a colder lower part of the lime kiln (see Figures 8

and 9). The colder lower part of the kiln is largely due to cooling air flowing through cooling zone without lime discharging.

Both adding extra COG and delaying the start of discharging lime after a stop in order to compensate for the heat lost during a stop seems to be beneficial to minimise variations in %CO_2 in lime. With different length of stop and operating conditions the amount of extra energy must be adjusted to correspond to the actual heat loss.

6.1 Limitations in the Modelling

One part of the model that lacks detailed information is flame modelling. Currently, the flame length and its profile for heat release is needed as input to the model. The longer the flame length, the more it extends into the cross-over channel, and consequently the more the temperature in the channel varies (Piringer 2015). To improve the model, the properties of the flame should be investigated further.

Another part of the modelling that needs manual input is the temperature decrease during an operation stop. A temperature profile of the lime kiln based on process data on channel temperature is assumed when starting lime kiln operation after a stop. Model development together with extensive analysis of process data and probably some tests must be done to enable good simulation of temperature change during a stop. Consequently, it seems that the sim-
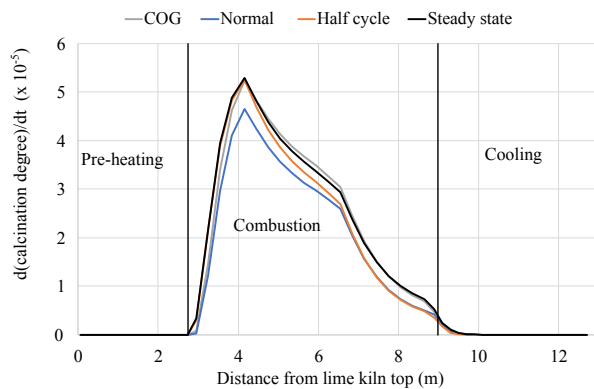
Figure 8. The average rate of calcination in the lime kiln for steady state and the three first cycles for the "Normal", "COG", and "Half cycle" cases.
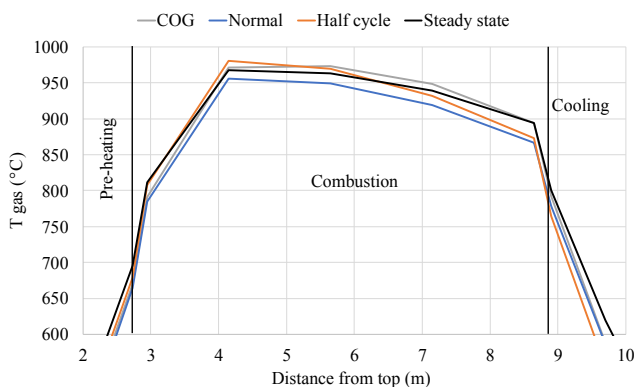
Figure 9. Average gas temperatures for steady state and the three first cycles for the "Normal", "COG", and "Half cycle" cases.

plest way is to assume a temperature loss and a temperature profile of the lime kiln depending on stop time.

7 Conclusion

We have developed an OpenModelica model of the PFR lime kilns at SSAB in Raahe. The model is a dynamic simulation model describing the calcination process of limestone in the lime kiln. The purpose of the model is to help in formulation of operations strategies and to aid in selecting important parameters to measure and monitor.

We have used the model to simulate a situation with a production stop for three different approaches on how to restart the kiln after the stop. The results from these simulations show that the model can indeed be useful for deciding operation strategies.

References

Agnieszka Bes. *Dynamic Process Simulation of Limestone Calcination in Normal Shaft Kilns*. PhD thesis, Otto-von-Guericke-Universität, Magdeburg, 2006.

T. Bluhm-Drenhaus, E. Simsek, S. Wirtz, and V. Scherer. A coupled fluid dynamic-discrete element simulation of heat and mass transfer in a lime shaft kiln. *Chemical Engineering Science*, 65(9):2821 – 2834, 2010. ISSN 0009-2509. doi:10.1016/j.ces.2010.01.015.

Duc Hai Do. *Simulation of Lime Calcination in Normal Shaft and Parallel Flow Regenerative Kilns*. PhD thesis, Otto-von-Guericke-Universität, Magdeburg, 2012.

F. Herz, B. Hallak, E. Specht, R. Gröpler, and G. Warnecke. Simulation of the limestone calcination in normal shaft kilns. In *10th European Conference on Industrial Furnaces and Boilers (INFUB)*, 7–10 April 2015.

Bastian Krause, Birk Liedmann, Jens Wiese, Patrick Buchera, Siegmar Wirtz, Hannes Piringer, and Viktor Scherer. 3D-DEM-CFD simulation of heat and mass transfer, gas combustion and calcination in an intermittent operating lime shaft kiln. *International Journal of Thermal Sciences*, 117: 121–135, 2017. doi:10.1016/j.ijthermalsci.2017.03.017.

Sanjeev Manocha and François Ponchon. Management of lime in steel. *Metals*, 8:686, 08 2018. doi:10.3390/met8090686.

Hannes Piringer. Key factors to minimise emissions from parallel flow regenerative lime kilns. Technical Report 11, ZKG International, 2015.

Hannes Piringer. Lime shaft kilns. *Energy Procedia*, 120:75–95, 2017. doi:10.1016/j.egypro.2017.07.156.

Andrej Senegačnik, Janez Oman, and Brane Širok. Analysis of calcination parameters and the temperature profile in an annular shaft kiln. part 1: Theoretical survey. *Applied Thermal Engineering*, 27(8):1467–1472, 2007. ISSN 1359-4311. doi:10.1016/j.applthermaleng.2006.10.001.

World Steel Association. Fact sheet — steel and raw materials, 2019. URL https://www.worldsteel.org/en/dam/jcr:f07b864c-908e-4229-9f92-669f1c3abf4c/fact_energy_2019.pdf.

Multivariate linear regression model of paste thickener

Jari Ruuska[1] Eemeli Ruhanen[2] Janne Kauppi[3] Sakari Kauvosaari[3] Mika Kosonen[3]

[1]Control Engineering, University of Oulu, Finland, `jari.ruuska@oulu.fi`
[2]Yara Suomi Oy, Siilinjärvi Mill, Finland
[3]Outotec (Finland) Oy, Finland

Abstract

The world is using a lot of materials in the day-to-day life. This requires a lot of mining of the ores from the ground. As the amount of ore is remarkable, also the amount of waste rock and for that reason the amount of the tailings is huge. The water content of the tailings is a subject to decrease. One potential technology for that is a paste thickener. In this paper, a multivariate linear regression model using paste line pressure difference as an output variable is described. The model can be utilized for the development of a new control strategy. Another model was formed using rake torque as an output variable.

Keywords: paste thickener, multivariate linear regression, pressure difference, rake torque

1 Introduction

Modern world relies on materials. Saying: "What you can't grow, you need to dig" still stands meaning that people need different kind of materials to make the products and the necessities for their everyday life. Caused by this fact, mining is a very important industry branch. Due to that, it needs to be made as efficient as possible. Together with the increase of the tonnage of the rock crushed, the amount of the waste rock also increases. The tailings from a flotation circuit contain a remarkable amount of water and if the water content can be lowered, the required area of tailings pond can be decreased. Due to constantly, more and more strict getting environmental permissions make this kind of development favourable. One way to do this is to add a paste thickener into the process chain. Thickening is a most economical method within the several dewatering techniques. The thickening process normally occurs in the large-diameter tanks where the solids particles settle under the influence of gravity, i.e. sedimentation. Adiguzel and Bascetin (2019) study the flow behaviour of the paste tailings, which consists of different pulp densities, was studied to define the effect of the particle size distribution of the tailings on a surface paste disposal. So to speak, the rheological behaviour of the tailings was examined. Jewell *et al.* (2002) raise a question into the discussion: the operation of the paste thickeners has different practical problems. One of

them is a narrow operating window as a small change in the underflow solids concentration may have a major impact on the properties (for example pumpability and flowability) of the underflow. Due to these issues, the control of the paste thickener can be challenging. The out coming paste is from time to time challenging to operate. At the worst, it causes a clogging in the process. This causes more need for the maintenance and even downtime. To be able to avoid this scenario, the process needs to be known better and the control to be more derivative. Bürger *et al.* (2005) have formulated a mathematical model based on the balance equations to simulate the dynamics. Betancourt *et al.* (2014) present a one-dimensional model based on the two nonlinear partial differential equations. Gheshlaghi *et al.* (2013) introduce a model based on the computational fluid dynamics and the population balance models. Chai *et al.* (2016) propose a controller including an underflow slurry flowrate (USF) pre-setting unit via a one-step optimal proportional and integral (PI) control with the un-modelled dynamics compensation, a fuzzy reasoning based USF set-point compensator and a switching mechanism using rule based reasoning. Setiawan *et al.* (2013) introduce a model-based control (MPC), which shows a significant improvement to controllability in the simulations. Tan *et al.* (2015) propose a model and Kalman filter used to make a coal parameter estimation showing an improved control result on the underflow solids concentration. Tan *et al.* (2016) present a MPC with the changing dynamics to handle fast, moderate and slow variables using the changing optimization horizon. Tan *et al.* (2017) introduce an additional rake torque constraint model to an earlier developed MPC to regulate further the underflow solids concentration and to prevent the rake liftings and the pumping problems. Tippett *et al.* (2014) propose a MPC with a changing sampling rate; the nearer steps had a shorter sampling rate than those in the distant future. This kind of an approach gives a possibility to fine-tune the control trajectory in the near horizon and still allows a long prediction and the control horizons for the slow dynamics. Xu *et al.* (2015) discuss a control strategy based on a mass balance model and the expert rules.

A multivariate linear regression model using paste line pressure difference as an output variable is presented.

2 Materials and methods

2.1 Process description

Only EU-area's phosphate mine, which is operated by Yara, is located at Siilinjärvi, Finland. The product of the ore extracted from the open pits is apatite concentrate, from which the fertilizers and phosphoric acid are the final products. The total ore mining annually is 11 Mt and the production of the apatite concentrate is 1 Mt.

The solids content of slurry increases typically in the thickeners. A clean overflow and the maximum solids concentration in the underflow are the general objectives. The agglomerates of the solids are formed by using the flocculants to increase the settling rate and thus to improve the overflow clarity. A very high on-line availability is typical for the thickeners' operation. Many industry branches use these vessels.

A thickener used in the mineral processing is introduced in Figure 1. The flocculants rate and the underflow rate are the two independent variables, which are typically used for the control. The feed rate is only used in an emergency situation to avoid the disturbances to the plant production. The dependent variables are for example rake torque, underflow density, overflow turbidity, solids interface level (bed depth), solids inventory (bed pressure), solid settling rate and underflow viscosity. (Bergh *et al.*, 2015)
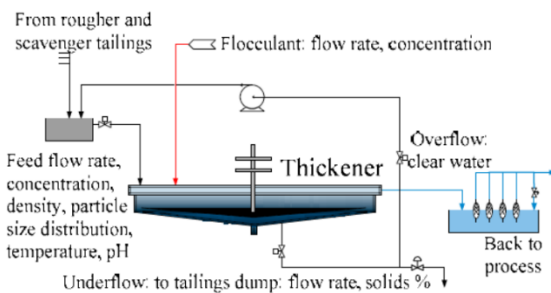
Figure 1. Typical thickener used in mineral processing

2.2 Process data

The data set used for the data-analysis contains data from Yara's both paste thickeners, but after discussing with the plant experts the data-analysis was limited only to the thickener one. The whole data set contained altogether 73 measurements from February -17 to February -18. The frequency of the measurements was one minute. An MPC-controller to reduce the underflow density variation is presented in Kosonen *et al.* (2017).

As the process data was used, a normal pre-processing procedure was needed. The rows containing NaN (not a number) were removed. The periods, when the values of the pressure difference of paste line (i.e. response variable) were not changing, were removed. Four feed and three thickener 1 measurements were delayed eight hours. The delay was defined by the plant experts. The definition was based on their process knowledge. The size of the data set was 265 000 rows; from which 130 000 was chosen for the training and 135 000 for the testing. The amount of the variables chosen for the modelling was sixteen; the selection was based on the data-analysis (reported elsewhere) and the plant expert knowledge.

2.3 Modelling

Multivariate linear regression (MLR) model is given by

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_M x_M = b_0 + \sum_{j=1}^{M} b_j x_j \quad , \quad (1)$$

where y is the output variable that is predicted, x_j are the input variables and b_j are the unknown model parameters. As the name indicates, MLR models are able to map the linear interactions variables. The model may also include the interaction and quadratic terms. A more detailed description on MLR can be found for example from Rencher and Christensen (2012).

2.4 Variable reduction

The variable reduction is done based on the statistics using p-value as an indicator. The smaller the p-value, the higher the significance of the variable is for the result. The reduction of the variables is executed so that after the first modelling round, the variable with the biggest p-value is removed and then the modelling is done again and the next variable is removed based on p-value. The procedure is repeated as long as the modelling result is improving.

2.5 Modelling objective

The modelling objective for this study was to model the paste thickener process using paste line pressure difference as an output variable. The changes in this variable give good information of the status of process and the material in the thickener. The rise in the pressure difference often is an indicator that the problems in the controllability of the process will occur or at least the probability for them is arising. To be able to predict this, a good model of a process is needed. Another model was formed using the same, reduced variable set as for the first model; the output variable was rake torque.

2.6 Static and dynamic model

The basic multivariate linear regression model is called a static model. The static model considering the average error of the previous ten data points is called a dynamic model.

3 Results and discussion

3.1 Paste line pressure difference model

The first modelling result for the training data with the sixteen variables is presented in Figure 2. As it can be seen, the variation is quite remarkable.
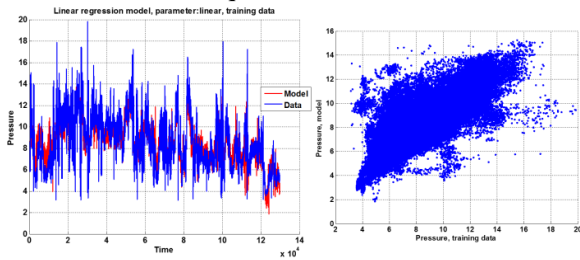
Figure 2. Model result for training data with sixteen variables

The test figures for the models with sixteen and six variables are presented in Figure 3. The variation of the six variables model is a bit smaller and it can follow the changes in the process circumstances a bit better. Still the variation is remarkable.
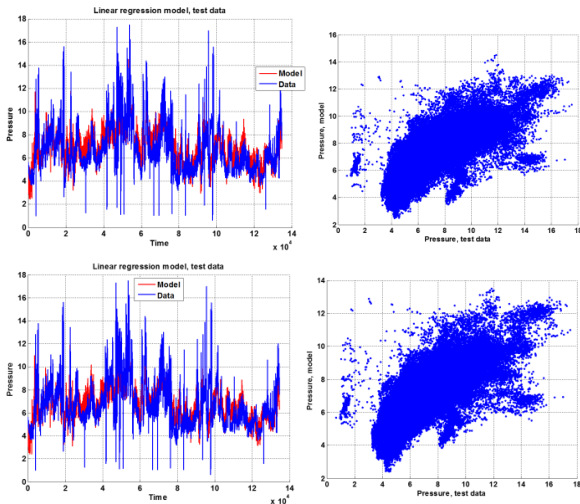
Figure 3. Test figures with 16 and 6 variables.

For the six variables model two further versions was done. The interaction and quadratic terms were allowed in those models. The test figures are presented in Figure 4. The variations are remarkable.
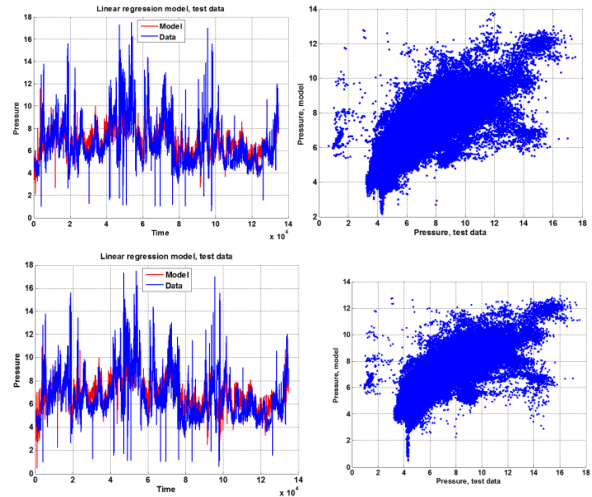
Figure 4. Test figures of six variable models with interaction and quadratic terms

The modelling error summary for the different models is given in Table 1.

Table 1. Modelling error summary

Model	Average error, training	Average error, test
16 variables	1,0971	1,0518
13 variables	1,1091	1,0597
12 variables	1,1102	1,0623
11 variables	1,1099	1,0574
10 variables	1,1084	1,0157
9 variables	1,1157	1,0041
8 variables	1,1150	1,0062
7 variables	1,1190	1,0082

As the modelling result didn't improve remarkably by reducing the variables, it was considered that something else should be done. As a one possibility, it was seen that the short term changes in the process are not captured by the current model structure and it could be handled by using a simple ten previous data points' error average addition to the model result. The five variables model test figures for a static and a dynamic model are presented in Figure 5. The model performance is improved quite considerably. The performance can be found in Table 2.
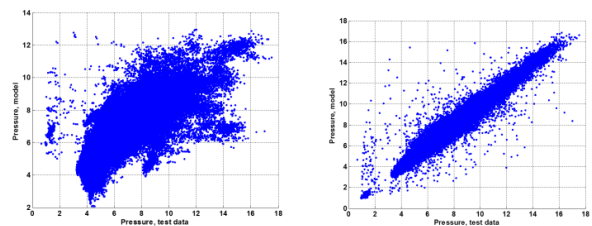
Figure 5. Test figures of static and dynamic model

Table 2. Comparison of a static and a dynamic model

Model	Error, training	Error, test
5 variables_static	1,1512	1,0012
5variables_dynamic	0,3186	0,2667

3.2 Rake torque model

Another model was formed using rake torque as an output variable. The model has four variables and the model result of the training data can be seen in Figure 6. The figure of the test data is presented in Figure 7. The figures are presented from the dynamic models. The performance of the models can be found in Table 3.
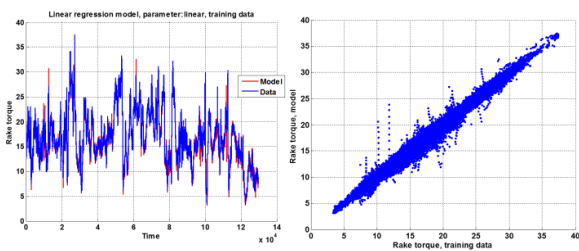
Figure 6. Model result of training data for dynamic model
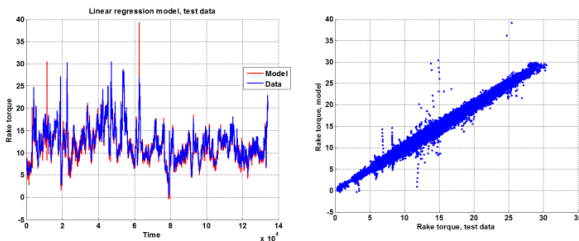
Figure 7. Test figures of dynamic model

Table 3. Comparison of static and dynamic model

Model	Error, training	Error, test
Rake torque	3,8441	3,9536
Rake torque_dynamic	0,3361	0,3364

Both models give satisfactory results, but there is still plenty room for the improvement. Therefore, other variables as an output variable should be tested, for example, the product of paste line pressure difference and rake torque or other combinations. Another modelling method should also be considered.

4   Conclusions

As the environmental permissions are getting stricter, it is important to make the processes in use as effective as possible. In this study, the paste thickener process was modelled using MLR-method. The modelling result was satisfactory. As a result from this study, the significant variables effecting on the paste line pressure difference were identified. This information can be utilized as an input for developing a new control strategy. It was still noticed that not all the needed variables were available for the modelling purposes. One of the phenomena, which are difficult to measure, is the changes in the rheology. It would be very useful to have a reliable measurement for it. The changes in the rheology can be quick and they have a significant effect on the pumpability of the paste. That kind of a measurement would be very useful on many different branches of the industry.

References

D. Adiguzel, A. Bascetin. The investigation of effect of particle size distribution on flow behavior of paste tailings *Journal of Environmental Management*, 243, 393-401, 2019. doi: 10.1016/j.jenvman.2019.05.039

L. Bergh, P. Ojeda and L. Torres. Expert control tuning of an industrial thickener, *IFAC-PapersOnLine,* 48, 86-91, 2015. doi: 10.1016/j.ifacol.2015.10.083

F. Betancourt, R. Bürger, S. Diehl and S. Farås. Modeling and controlling clarifier–thickeners fed by suspensions with time-dependent properties, *Minerals Engineering*, 62, 91-101, 2014. doi: 10.1016/j.mineng.2013.12.011

R. Bürger, K.H. Karlsen and J.D. Towers. Mathematical model and numerical simulation of the dynamics of flocculated suspensions in clarifier–thickeners, *Chemical Engineering Journal*, 111, 119-134, 2005. doi: 10.1016/j.cej.2005.02.007

T. Chai, Y. Jia, H. Li and H. Wang, H. An intelligent switching control for a mixed separation thickener process, *Control Engineering Practice*, 57, 61-71, 2016. doi: 10.1016/j.conengprac.2016.07.007

M.E. Gheshlaghi, A.S. Goharrizi, A.A. Shahrivar and H. Abdollahi. Modeling industrial thickener using computational fluid dynamics (CFD), a case study: Tailing thickener in the Sarcheshmeh copper mine, *International Journal of Mining Science and Technology,* 23, 885-892, 2013. doi: 10.1016/j.ijmst.2013.11.002

R.J. Jewell, A.B. Fourie and E.R. Lord. *Paste and Thickened Tailings: A Guide.* Australian Centre for Geomechanics, University of Western Australia, 2002.

M. Kosonen, S. Kauvosaari, S. Gao and B. Henriksson. Performance optimization of paste thickening, in A Wu & R Jewell (eds), *Proceedings of the 20th International Seminar on Paste and Thickened Tailings*, University of Science and Technology Beijing, Beijing, China, 13-22, 2017. doi: 10.36487/ACG_rep/1752_02_Kosonen

A.C. Rencher and W.F. Christensen (2012) "Chapter 10, Multivariate regression – Section 10.1, Introduction", *Methods of Multivariate Analysis*, Wiley Series in Probability and Statistics, 709 (3rd ed.), John Wiley & Sons. ISBN: 978-0-470-17896-6

R. Setiawan, C.K. Tan, J. Bao and G. Bickert. Model
Predictive Control of a Paste Thickener in Coal
Handling and Preparation Plants, *10th IFAC
International Symposium on Dynamics and Control of
Process Systems*, Mumbai, India, 247-252, 2013. doi:
10.3182/20131218-3-IN-2045.00123

C.K. Tan, R. Setiawan, J. Bao and G. Bickert, G. Studies on
parameter estimation and model predictive control of
paste thickeners, *Journal of Process Control*, 28, 1-8,
2015. doi: 10.1016/j.jprocont.2015.02.002

C.K. Tan, M.J. Tippett and J. Bao. Model predictive control
with non-uniformly spaced optimization horizon for
multi-timescale processes, *Computers and Chemical
Engineering*, 84, 162-170, 2016. doi:
10.1016/j.compchemeng.2015.08.010

C.K. Tan, J. Bao and G. Bickert. A study on model
predictive control in paste thickeners with rake torque
constraint, *Minerals Engineering*, 105, 52-62, 2017. doi:
10.1016/j.mineng.2017.01.011

M.J. Tippett, C.K. Tan and J. Bao. Non-Constant Prediction-
Step MPC for Processes with Multi-Scale Dynamics,
19th IFAC World Congress, Cape Town, South Africa,
3068-3073, 2014. doi: 10.3182/20140824-6-ZA-
1003.01093

N. Xu, X. Wang, J. Zhou, Q. Wang, W. Fang and X. Peng.
An intelligent control strategy for thickening process,
International Journal of Mineral Processing, 142, 56-62,
2015. doi: 10.1016/j.minpro.2015.01.007

Predicting Centrifugal Compressor Off-design Operation in an Operator Training Simulator

J. Fredrik Nyholm

Neste Engineering Solutions, Finland, `fredrik.nyholm@neste.com`

Abstract

Operator training simulators (*OTS*) are a valuable tool for improving the competence of both process operators and engineers. As pumps, compressors and turbines are vital parts of any process plant, accurate modeling of these machines is crucial for a dynamic process simulator. The behavior of the machines is usually described by operating curves provided by the manufacturer. However, these curves are strictly valid only for the specific process conditions they were created for. Accurate prediction of the operation outside of these specific conditions may present a formidable challenge. To further improve the accuracy the NAPCON Simulator, a new dynamic modeling method was developed for the centrifugal compressors. The stability of the developed model was verified by dynamic ramp tests of the operating conditions. Furthermore, the accuracy of the developed model was determined by performing process data tests. The results were promising and, in most cases, an improvement from the performance of the currently implemented model.

Keywords: dynamic simulation, operator training simulator, centrifugal compressors, off-design operation, exit flow coefficient

1 Introduction

Main features of a training simulator include sophisticated models of all process units with fast simulation of various operating states and fault scenarios. A well-designed OTS allows extensive and customizable training functionality. High-fidelity process models that predict accurate thermal performance and chemical conversion for complex process equipment ensure realistic process response. The process model is complemented by automation models containing process controls, safety functions and sequence loops. This all is presented to the trainee by a comprehensively emulated DCS interface complete with trends and alarms, providing a realistic process control experience. (Qvist *et al*, 2018)

NAPCON Simulator is based on an in-house software called ProsDS, which is programmed in ANSI Common Lisp. The simulator features a library of all commonly used process equipment. The high-fidelity process feedback is achieved by basing the simulation on physical phenomena and chemical reactions. The process equipment is connected by advanced piping network modeling with two-phase flow calculations. The process is modeled on hierarchical flowsheets. Equipment states as well as connectivity can be examined even during active simulation. Calculation steps in the process and automation models of less than one second provide accurate results of the dynamic simulation. The information both inside as well as between the process units is updated at each simulation interval.

In ProsDS the pumps are modeled using theoretical maximum head at zero flow and pressure loss due to flow. The head H_{max} can be given either as pressure difference or height of fluid column. The pressure loss is modeled using the equation for the valve flow coefficient CV (m^3*s^{-1}*Pa$^{-0.5}$), which relates the pressure loss Δp (Pa) to the volumetric flow $\dot{V}$ (m^3/s) as a quadratic function. The CV value is calculated with Equation 1, where SG (-) is the specific gravity of the pumped fluid. The operating curve for the pump is thus given by Equation 2. A curve of this sort is usually accurate enough for pumps around their normal operating points. Changes in the rotating speed affect the head of the pump quadratically. (Nyholm, 2020)

$$CV = \dot{V} \sqrt{\frac{SG}{\Delta p}} \qquad (1)$$

$$H = H_{max} - SG \frac{\dot{V}^2}{CV^2} \qquad (2)$$

The compressors are currently modeled as pumps with some extra features. For compressor curve fitting a correction function is used, which takes the actual load of the compressor into account. A correction factor for the effect of guide vanes is also available. Functions for these correction factors are modeled individually for each compressor by the person doing the modeling work. Any other calculations regarding the operating curves are usually carried out as external calculations modeled outside of the machine. No automatic curve correction for changes in e.g. molecular weight or inlet temperature exists in the simulator, these are defined by the modeler as external calculations. For compressors in general, the curve type should be steeper at higher loads

than what the *CV*-based quadratic curve is. (Nyholm, 2020)

The power requirement of the machines is also modeled in a simple manner. It is based on the flow and pressure difference over the machine, with a corresponding efficiency. The outlet temperature is also modeled in a simple way, based on either an empirically approximated constant or on the inlet and outlet pressures. (Nyholm, 2020)

2 Compressor theory

The centrifugal compressor consists of an impeller inside a casing, an inlet leading the gas to the impeller and an exit leading out from the volute surrounding the impeller. The impeller may also be surrounded by a vaned diffuser, but the diffusers in the machines used in the oil refining and petrochemical industries usually come without vanes. If the machine consists of multiple compression stages, the gas is led from the exit, through deswirling vanes, into the inlet of the next impeller. The energy of the gas is increased by increasing its speed, which is done by the impeller. The diffuser and volute then convert the speed to a pressure and temperature increase by efficiently reducing the fluid velocity.

The compressor is characterized by its performance map, where the produced head (or discharge pressure), efficiency and required power are plotted against the inlet volume flow. The manufacturer usually provides a number of performance maps valid for varying inlet conditions. The compressor has an operating range defined between the surge point and choke point.

Off-design operation of centrifugal pumps based on a performance curve can typically be described with reasonable accuracy by the affinity laws. (Pettersson *et al*, 1990). Even if centrifugal compressors and pumps share many similarities both in operating principles and construction, the compressors are more complex machines and the affinity laws are not sufficient to accurately describe a large operating regime for them.

In dynamic simulation the machine is subjected to a wide range of operating conditions and, thus, the operating curve of the machine should be scaled smoothly in a user-friendly manner to match the governing conditions, while maintaining a sufficient accuracy of the predicted performance. As directly scaling the performance curves with the affinity laws is not viable for compressors, approaches that consider the vital machine and operating condition parameters influencing the compressor performance are needed for an accurate prediction of new performance curves. One option is the computationally heavy CFD simulation of the machine. Another approach would be to utilize empirical correlations for the predictions. However, both methods require detailed information of the machine geometry, which usually is hard to come by.

Converting the performance curve to a dimensionless form, from which performance curves at new operating conditions can be predicted, appears to be the most viable option for an accurate implementation in an OTS. Using this method, the head is converted into the head coefficient ψ (-). For the volume flow either the inlet flow coefficient φ_1 (-) or the exit flow coefficient φ_3 (-) is used. Using the inlet flow coefficient, an array of dimensionless curves is obtained and, as the impeller tip speed Mach number varies, the needed dimensionless curve can be interpolated or extrapolated from the curve array and the converted back to the sought performance curve in natural scale. The aim of utilizing the exit flow coefficient is to obtain a single dimensionless performance curve for the compressor, which can then be used to calculate all the curves in natural scale. The advantage with both methods is that only the manufacturer performance curves, impeller diameter and blade tip height and knowledge of the governing process conditions are needed for the predictions. As both the impeller diameter and blade tip height are typically provided in the datasheet of the machine, the difficulties of obtaining information of the detailed geometry of the machine can be avoided.

The presented inlet flow coefficient method requires the impeller diameter and the performance maps of the manufacturer, as well as the reference conditions for these, for transforming the curves to the dimensionless variant. The governing process conditions are needed for estimation of the new performance curves. The impeller tip speed Mach number M_u (-) can be calculated by Equation 3, where U (m/s) is the blade tip speed, given by Equation 4. The impeller diameter d (m) and angular velocity ω (1/s) are used for calculating U. Additionally, the isentropic exponent κ (-), compressibility Z (-), temperature T (K), universal gas constant R (J*kmol^{-1}*K^{-1}) and gas molecular weight MW (kg/kmol) are needed for calculating the Mach number. The index 1 refers to the compressor inlet. If two sets of inlet conditions have equivalent Mach numbers, their dimensionless performance curves will also be the same.

$$M_u = \frac{U}{\sqrt{\kappa_1 Z_1 T_1 \frac{R}{MW}}} \qquad (3)$$

$$U = \pi d \omega \qquad (4)$$

Conversion of the performance curves themselves is done by transforming the polytropic head into the dimensionless head coefficient ψ and the volumetric flow into the inlet flow coefficient φ_1, as described by Equations 5 and 6, where H_p (J/kg) is the polytropic head.

$$\psi = \frac{2H_p}{U^2} \qquad (5)$$

$$\varphi_1 = \frac{4\dot{V}}{\pi d^2 U} \qquad (6)$$

After the conversion, an array of dimensionless operating curves, each corresponding to a certain Mach

number, has been obtained. Using Equation 3, a new Mach number $M_{u,new}$ is calculated from the new inlet conditions. Based on the relation of the new Mach number to the Mach numbers of the previously calculated dimensionless curves, the full dimensionless curve, consisting of the new head coefficient ψ_{new} and the new inlet flow coefficient $\varphi_{1,new}$, is created by a suitable curve approximation method. Lastly, the actual new operating curve, consisting of $H_{p,new}$ plotted against $\dot{V}_{new}$, is calculated from the new dimensionless curve using Equations 7 and 8.

$$H_{p,new} = \psi_{new} \frac{U_{new}^2}{2} \qquad (7)$$

$$\dot{V}_{new} = \varphi_{1,new} U_{new} d^2 \frac{\pi}{4} \qquad (8)$$

When operating within the Mach numbers given in the design data, using this method the new polytropic head curves can be approximated within an accuracy of 3%. The method does, however, have some weaknesses. It is strictly valid only if each stage of the compressor is treated separately. In addition, the accuracy of the prediction quickly deteriorates if the new dimensionless curves are extrapolated from those derived from the manufacturer data. (Ravi *et al*, 2015)

If the head coefficient ψ is used in conjunction with the exit flow coefficient φ_3 instead of the inlet flow coefficient φ_1, the performance of each compressor stage can be presented in a single dimensionless curve for all conditions up to a critical Mach number. For Mach numbers above the critical values, the curves will be reduced and lie to the left of the curve for the lower Mach numbers. During tests performed for two compressors, both the head coefficients and efficiencies could be presented in a single curve while operating below the critical Mach numbers, even as molecular weights of the compressed gas ranged from 29 g/mol to 365 g/mol. A central finding is that a single gas parameter does not affect the flow similarity. Rather, the combination of the parameters making up the tip speed Mach number is what counts as a flow similarity characteristic. (Sheets, 1952; Lüdtke, 2004)

To calculate φ_3 Equation 9 is used, where b_2 (m) is the impeller blade tip height. Indices *2* and *3* refer to the impeller exit and compressor exit, respectively. The impeller diameter and blade tip height are usually provided in the manufacturer data sheets. The ratio of the static volumetric flow rate $\dot{V}_3$ to the inlet volumetric flow rate is given by Equation 10, from which $\dot{V}_3$ is easily solved. The ratio of the temperatures can be solved from Equation 11, where m (-) is the polytropic temperature exponent. (Liese, 2017; Lüdtke, 2004; Sinnot 2005)

$$\varphi_3 = \frac{\dot{V}_3}{\pi d b_2 U} \qquad (9)$$

$$\frac{\dot{V}_3}{\dot{V}_1} = \frac{T_3}{T_1} \frac{p_1}{p_3} \frac{Z_3}{Z_1} \qquad (10)$$

$$T_3 = T_1 \left(\frac{p_3}{p_1}\right)^m \qquad (11)$$

Once the dimensionless curve consisting of ψ and φ_3 has been constructed, the performance curve for the new operating conditions can be calculated. Equation 7 is used to calculate the new values for the polytropic head. The values for the inlet volume flow at the new process conditions can be calculated with Equation 12, where $\dot{V}_3$ is given by Equation 13. (Nyholm, 2020)

$$\dot{V}_1 = \dot{V}_3 \frac{T_1}{T_3} \frac{p_3}{p_1} \frac{Z_1}{Z_3} \qquad (12)$$

$$\dot{V}_3 = \varphi_3 \pi d b_2 U \qquad (13)$$

The new temperature ratio is given by Equation 14 and the new outlet pressure p_3 (Pa) is given by Equation 15. The polytropic index n (-) can be calculated from κ and the polytropic efficiency η_p (-) according to Equation 16. The value of the polytropic efficiency should correspond to the value of φ_3. Equation 17 is used in the calculation of the compressibilities, where V_m (m³/mol) is the molar volume. (Nyholm, 2020)

$$\frac{T_1}{T_3} = \frac{1}{\left(\frac{p_3}{p_1}\right)^m} \qquad (14)$$

$$p_3 = p_1 \times \sqrt[\left(\frac{n-1}{n}\right)]{\frac{H_p}{\frac{R}{MW} Z_{avg} T_1 \left(\frac{n}{n-1}\right)} + 1} \qquad (15)$$

$$n = \frac{1}{1 - \frac{\kappa - 1}{\kappa \eta_p}} \qquad (16)$$

$$Z = \frac{p V_m}{RT} \qquad (17)$$

Once H_p and $\dot{V}_1$ at the new process conditions are calculated for the range of the dimensionless curve, the new operating curve is ready for use. Unfortunately, even this method is strictly applicable only for a single compression stage. A general way to describe the operation of the whole compressor in a single curve would be a great benefit for accurate dynamic simulation, but alas no such thing as an invariant performance curve seems to exist for multistage compressors. (Lüdtke, 2004)

The outlet temperature of the compressor can be calculated with Equation 11. When not operating at critical conditions, m reduces to Equation 18. (Sinnot, 2005)

$$m = \frac{\kappa - 1}{\kappa \eta_p} \qquad (18)$$

For a simple approximation of the compressor power consumption P (W), Equation 19 can be used, where h (J/kg) is the specific enthalpy of the gas, $\dot{m}$ (kg/s) is the mass flow and η (-) is the general efficiency of the compressor. (Westerlund, 2009)

$$P = \frac{\dot{m}(h_3 - h_1)}{\eta} \qquad (19)$$

3 Developed dynamic model

For a dynamic simulator used in operator training, correct and accurate behavior is vital. For a single piece of process equipment in the model, it is of central importance that both the direction and size of the dynamic response to changes in the process conditions match reality. Bearing in mind that individual process equipment are an integral part of a larger process model also accentuates the importance of the accuracy of the absolute values for e.g. flow and temperature calculated by the model, as these affect the downstream process model. As long as the dynamic behavior of the model is correct, it may be possible to use empiric correction factors to adjust the absolute process values to correct levels. However, excessive use of such corrections are likely to cause further model issues sooner or later. False behavior of both individual equipment and the total process model are severely detrimental for use of the model in both educational and engineering applications.

The simulator should, at the bare minimum, be able to run in real-time, which limits the acceptable computational complexity for any single piece of machinery or equipment in the process model. As the process model needs to be able to handle a plethora of operating conditions and upset scenarios, the compressor model should also be flexible and robust. Changes in the inlet temperature, gas composition, pressure and flowrate, rotation speed of the compressor, as well as changes in the outlet pressure, should affect the compressor operation in an appropriate manner.

As computationally heavy methods such as CFD modeling of the entire machine is not an option for dynamic simulation, the best way to accurately calculate the operation of the machine would seem to be to use the performance curves and determine how they change with varying operating conditions.

The compressor operation is characterized by the operating point, outlet temperature and consumed power, of which the operating point is most important. As such, the focus during development of the new model was mainly on calculating the operating point and describing its behavior during off-design operation.

Bearing the mentioned criteria in mind, the exit flow coefficient method appeared as the most suitable for operating curve prediction. Power and outlet temperature calculations were also implemented, but in a relatively simple manner.

The developed method starts off by determining the dimensionless operating curve for the compressor. Equations 5 and 9 were used to transform the manufacturer curves to dimensionless form. A simplified form of Equation 10 was used, without the compressibilities. Equations 11 and 15 were used for calculating the outlet pressure and temperature. In Equation 15 the inlet compressibility was used, instead of the average. The modeled compressor was modeled as a single-stage operation, so only one dimensionless curve was used for the entire machine. Calculation of m was done with Equation 20. To compensate for the reduced assumption made in Equation 18, the polytropic efficiency was added. The tuning parameter m_{corr} (-) was also added. This parameter is common for all the curves and can be used by the modeler to align them. (Nyholm, 2020)

$$m = \frac{n-1}{n\eta_p} m_{corr} \qquad (20)$$

Once polynomials have been fitted to the dimensionless performance curve and the dimensionless polytropic efficiency curve, the coefficients of the polynomials are given as input to the constructed compressor model in ProsDS. The model reads the temperature, pressure, molar weight, specific heat capacity, molar flow, molar volume and enthalpy from the inlet flow. The pressure and enthalpy are read from the outlet stream. The dimensionless curve in ProsDS is defined as a set of points. The polytropic head and inlet volume flow are calculated with Equations 7 and 12. The model determines the operating point through interpolation with the help of the governing outlet pressure and writes the volume flow to the inlet stream. Using Equation 21 the compressor model then calculates and writes the temperature to the outlet stream. The compressibility was tested as a correction factor for the outlet temperature. Based on some tests in Excel, adding Z to the exponent seemed to improve the accuracy of the temperature calculation at design point operation. The effect of the correction depends on the used equation of state. The power consumption was calculated with Equation 22, where $\dot{\xi}$ (mol/s) denotes the molar flow and h (J/mol) is the molar enthalpy of the stream. (Nyholm, 2020)

$$T_3 = T_1 \left(\frac{p_3}{p_1}\right)^{(n-1)/nZ} \qquad (21)$$

$$P = \dot{\xi}(h_3 - h_1) \qquad (22)$$

The special operating conditions surge and choke were also considered in the model, although in a more indicative manner. The limits for surge and choke were taken as the end points from the manufacturer curves If the model predicted surge the flow would drop to zero, whereas if choke was predicted the flow would be limited at a maximum value. (Nyholm, 2020)

4 Results

To evaluate the stability, accuracy and dynamic behavior of the constructed model, a series of test were performed in the simulator. The results were compared to both manufacturer data and process data. Differences in the calculated values of κ and Z and the values given

by the manufacturer were noted. To determine the effect of these differences the model was tested at a number of operating points along a single operating curve. It was determined that the difference had a lowering effect on the predicted operating curve. Effects on the power consumption were also investigated. For this, both the ideal and Soave-Redlich-Kwong (*SRK*) equations of state were tested. The tests showed that the difference in the values increased the power consumption. If the values were given as constants matching the process data, the SRK equation of state gave better results. If the values were calculated by the simulator the ideal equation of state gave results closer to the manufacturer data. (Nyholm, 2020)

The model was also subjected to more dynamic tests to verify its stability and correct dynamic behavior. During these tests, the temperature, pressure and molecular weight of the inlet stream as well as the pressure of the outlet stream were adjusted by a ramp function in the simulator. Some issues could be observed when the values were ramped down to extremely low, abnormal levels, such as temperatures close to absolute zero or pressures close to complete vacuum. Around realistic operating regimes the model displayed realistic dynamic behavior and smooth transitions between surge, normal operation and choke. (Nyholm, 2020)

Once the validity of the concept had been proved in Excel and the stability of the constructed model ascertained with the dynamic tests, the testing could progress to assess the actual accuracy of the model by performing the process data tests. Process data of three different operating cases was available for the modeled compressor. The data of the stream inlet temperature, composition and pressure as well as outlet pressure, was fed to the test model, which then predicted the inlet volume flow, outlet temperature and compressor power consumption. The old compressor model implemented in ProsDS was tested in parallel to the newly developed model for comparison.

The first case to be tested was the shutdown operation. This case was tested twice; once with the ideal equation of state and once with the SRK equation of state. The predicted mass flow and temperature showed a smaller offset form the process data when using the ideal equation of state, although the power consumption had a somewhat larger offset. As the ideal equation overall gave better results it was used in the other tests as well.

During the first test case it was clearly visible that the mass flow predicted by the new model was an improvement. At conditions close to normal operation the predicted flow lies slightly above the measured flow. As the process conditions diverge from normal operation the offset between the curves can be seen to increase. The reason behind this would seem to be increasing inaccuracies in the predicted operating

curves, as the predictions for some of the curves will lie more to the left of the manufacturer performance curves when moving from the rated operating conditions towards conditions resembling the initial circulation case. The shape of the trend clearly followed the process data better than the predictions of the old model. The oscillations seen towards the end of the predicted curve are due to the model balancing on the verge of surge.
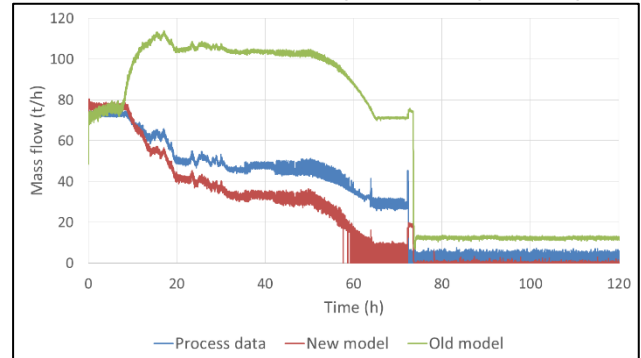
Figure 1. Test result for the suction mass flow of the shutdown case using the ideal equation of state.

As can be seen in Figure 2 the temperature calculation of the old model works quite well. The shape of the predicted curve is very similar to that of the process data. During normal operation, this prediction is slightly better than the new model's prediction, but as the conditions diverge from normal operation the new model gives a better prediction. The predicted curves have been cut as the machine is turned off, since they are no longer accurate after that point. (Nyholm, 2020)
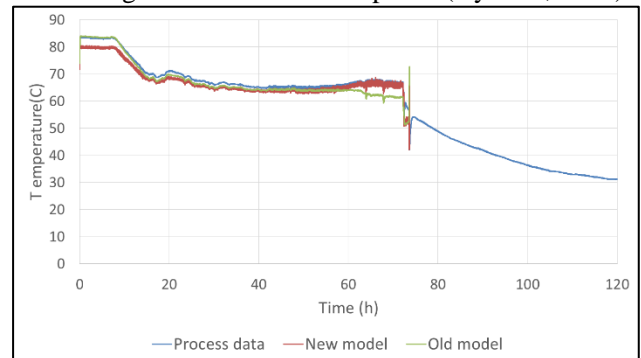
Figure 2. Test result for the outlet temperature of the shutdown case.

The predicted electric current initially lies above the process data with a bigger offset than in the SRK test case but approaches the process data curve towards the end of the operation, as seen in Figure 3. The predicted current of the old model differs completely from the process data. (Nyholm, 2020)
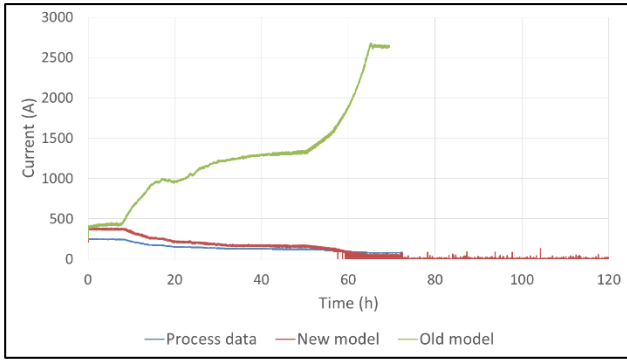
Figure 3. Test result for electric current of the shutdown case.

The test results of the predicted mass flow for the normal operation case and the startup case are shown in Figure 4 and Figure 5. The new model gives a clearly improved prediction compared to the old model, although some discrepancy is visible in the startup case, where the model is predicting surge for some time at the start. The predictions of the outlet temperature and electric current behaved in a similar manner as in the shutdown case. (Nyholm, 2020)
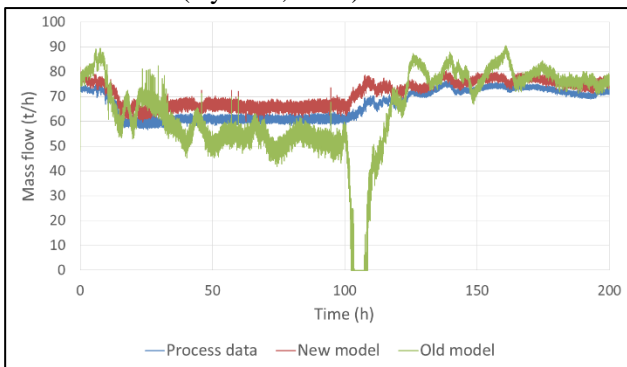
Figure 4. Test result for the suction mass flow of the normal operation case.
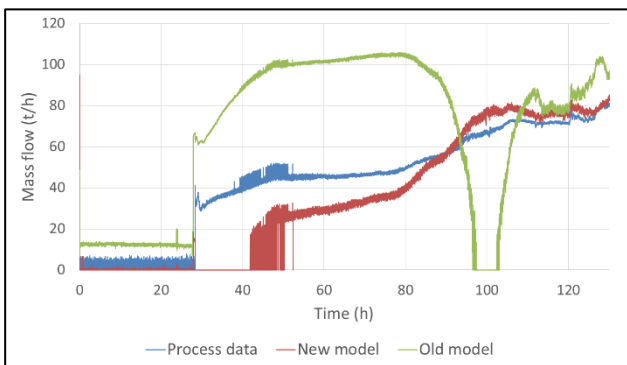
Figure 5. Test result for the suction mass flow of the startup case.

5 Discussion

5.1 Benefits

The benefits of an OTS during the design and commissioning phase of a project include functions for testing the design of control and safety automation systems and functionality of operator user interface layouts, virtual testing of process design and changes and planning of safe startups and shutdowns. When used for operator training, the operators can safely be trained for coping with any fault situation. Additionally they can be trained in how to operate the facility for maximal production efficiency. Increased availability and reduced downtime of the process plant are direct benefits. (Qvist *et al*, 2018)

The improved compressor model will benefit the NAPCON Simulator OTS by increasing the accuracy of the process models especially during shutdown and startup operations. A more accurate performance of the models during these operations is crucial for realistic response during training scenarios of such situations.

The new method also allows for a faster modeling of the machines, as the need for external calculations are reduced. The person doing the modeling will not have to search for and tune the correlations between the process parameters and compressor performance as extensively as previously, as they are already included in the model.

5.2 Future development

Based on the results of the various tests performed for the developed model several areas of improvement were identified. These subjects mainly relate to the operating curve, outlet temperature calculation and power calculation. Based on some preliminary tests, the most promising way to address the greater inaccuracies of the model, namely the offset of the flow prediction, appears to be implementing a method for scaling the polytropic efficiency to match the operating conditions. (Nyholm, 2020)

6 Conclusions

A new dynamic simulation model for centrifugal compressors was constructed using dimensionless operating curves. With the help of a tuning parameter the method could be used for a multi-stage compressor with promising results.

The developed model is stable over a wide range of operating conditions, supporting the feasibility of the selected method of modeling the compressor operation. When examining the tests and figures of section 4, it can be found safe to claim that modeling of the compressor flow and required electric current were substantially improved by the developed model.

The flow and outlet temperature calculations of the new model performed better when using the ideal

equation of state, while the electric current calculation fared better when the SRK equation of state was used. When using the SRK method the outlet temperature predicted by the developed model seemed better compared to the old model's prediction. When the ideal equation of state was used, however, the outlet temperature predicted by the old model described the process data more accurately for the most part of the test. The accuracy of the new model's temperature prediction was better only shortly before the machine was shut off or shortly after it was turned on.

When studying the graphs in section 4, the possibility of inaccuracies and various uncertainties pertaining to both the manufacturer data and the measured process data should be kept in mind. Albeit the developed model overall appears to be a clear improvement, it still suffers from certain shortcomings and requires further development to overcome these issues.

References

Eric Liese, Stephen E. Zitney. The Impeller Exit Flow Coefficient as a Performance Map Variable for Predicting Centrifugal Compressor Off-Design Operation Applied to a Supercritical CO2 Working Fluid. *ASME Turbo Expo 2017: Turbomachinery Technical Conference and Exposition*, Vol 9, 2017.

Klaus H. Lüdtke. *Process Centrifugal Compressors*. Springer-Verlag Berlin Heidelberg GmbH. 2004.

J. Fredrik Nyholm. *Improved modeling of rotary momentum mass transport equipment*, M.Sc. thesis, Åbo Akademi University, 2020.

Frank Pettersson, Hans Skrifvars, Kurt Lundqvist, Kaj Fagervik, Tapio Westerlund. *Ett övningsarbete i pumpdimensionering*. Åbo Akademi University. 1990.

Pekka Qvist, Ari Vepsäläinen and Prateek Agarwal. Develop competence in the process industry with a digitalized operator learning path. *Hydrocarbon Processing*, 97(7), 2018

Arjun Ravi, Lukasz Sznajder, Ian Bennett. Compressor map prediction tool. *IOP Conference Series: Materials Science and Engineering*, Vol 90, 2015. doi:10.1088/1757-899X/90/1/012042

H. E. Sheets. Nondimensional Compressor Performance for a Range of Mach Numbers and Molecular Weights. *Transactions of the ASME*, 74(1):93-102, 1952.

R. K. Sinnot. *Coulson & Richardson's Chemical Engineering Volume 6: Chemical Engineering Design*. Elsevier Butterworth-Heinemann. 2005.

Tapio Westerlund. *Anläggnings- och systemteknik*. Tekniska fakulteten vid Åbo Akademi. 2009.

Application of population balance equation for continuous granulation process in spherodizers and rotary drums

Ludmila Vesjolaja[1]   Bjørn Glemmestad[2]   Bernt Lie[1]

[1]Department of Electrical Engineering, IT and Cybernetics, University of South-Eastern Norway,
`{ludmila.vesjolaja,bernt.lie}@usn.no`
[2]Process Modeling and Control Department, Yara Technology Center, Norway, `bjorn.glemmestad@yara.com`

Abstract

In this paper, a dynamic model for a granulation process is developed. A population balance is used to capture dynamic particle size distribution in the spherodizer model and in the rotary drum granulator model. Particle growth due to layering is assumed in the spherodizer simulation model, while particle binary agglomeration is taken as the main granulation mechanism in the rotary drum simulation model. The developed models are 2-dimensional(2D) models that are discretized in terms of its internal coordinate (particle diameter), external coordinate (axial length of the granulators). Simulations using the developed models provide valuable data on dynamic fluctuations in the outlet particle size distribution for the granulators. In addition, the simulations results give a valuable information for the control studies of the granulation process. The simulation results showed that the extension of the model from 1D model to 2D model using the discretization of the external coordinate in the model, introduces a transport delay that is important in control studies.

Keywords: spherodizer, rotary drum, population balance, dynamic model, time delay

1 Introduction

Granulation processes are used in a wide range of industrial applications, including fertilizer industries. Fertilizer manufacturing using the granulation process has received considerable research interest during the last few decades, due to (i) the increasing requirements for efficient production of high quality fertilizers for increased food production in a growing global population, and (ii) difficult process control and operation, e.g., among others (Herce et al., 2017; Ramachandran et al., 2009; Valiulis and Simutis, 2009; Wang et al., 2006) and (Cameron et al., 2005) have focused their research on granulation processes. This paper is focused on the last part of the mineral fertilizer production, i.e. on the granulation loop. The granulation loop is used to produce different grades of fertilizers. A typical schematic of a granulation process with the recycle loop is shown in Figure 1. The granulation loop consists of a granulator, granule classifier (screener), and a crusher. The granulator receives the fines from the external particle feed, as well as from the recycled styream. These particle feeds are sprayed with a fertilizer liquid melt (slurry), and granules are formed. Different granulation mechanisms depending on the granulator type and conditions are responsible for these granule formation.

Process control of granulation loops is challenging. Typically, the PSD of the granules leaving the granulator is wider compared to the required PSD of the final product. A typical recycle ratio between the off-spec particles (80 %) and the required product-sized particles (20 %). Thus, it is important to develop a dynamic model that could further be used in control relevant studies. This paper is focused on developing dynamic models of two types of granulators, namely spherodizers and rotary drums. Depending on the granulator type and operating conditions, different granulation mechanisms (granule formation mechanism) are predominant. In spherodizers, the main granulation mechanism is particle growth due to layering. Layering is a continuous process during which particle growth occurs due to a successive coating of a liquid phase onto a granule. In rotary drum granulators, on the other hand, particle collision occurs, and thus particle agglomeration contributes significantly to particle size change. In this paper, binary particle agglomeration is assumed for population balance (PB) modeling. Binary agglomeration refers to a granule formation mechanism that occurs due to successful collision of two particles, resulting in the formation of a larger, composite particle (Litster and Ennis, 2004; Vesjolaja et al., 2018).

This paper is an extension of our previous study that is summarized in (Vesjolaja et al., 2018). Here, the dynamic model of the granulator is improved by increasing the dimensionality of the model, i.e., a 2D model instead of a 1D model is developed. The improved dynamic model provides valuable data on dynamic fluctuations in the outlet particle size distribution for the granulators. Thus, the contributions of this paper are: (i) the 1D model is extended to the 2D model (ii) developed 2D models are applied for two types of granulators, spherodizers and rotary drums, and (iii) two different numerical schemes, a finite volume scheme and a sectional scheme, are applied to the developed 2D dynamic models.
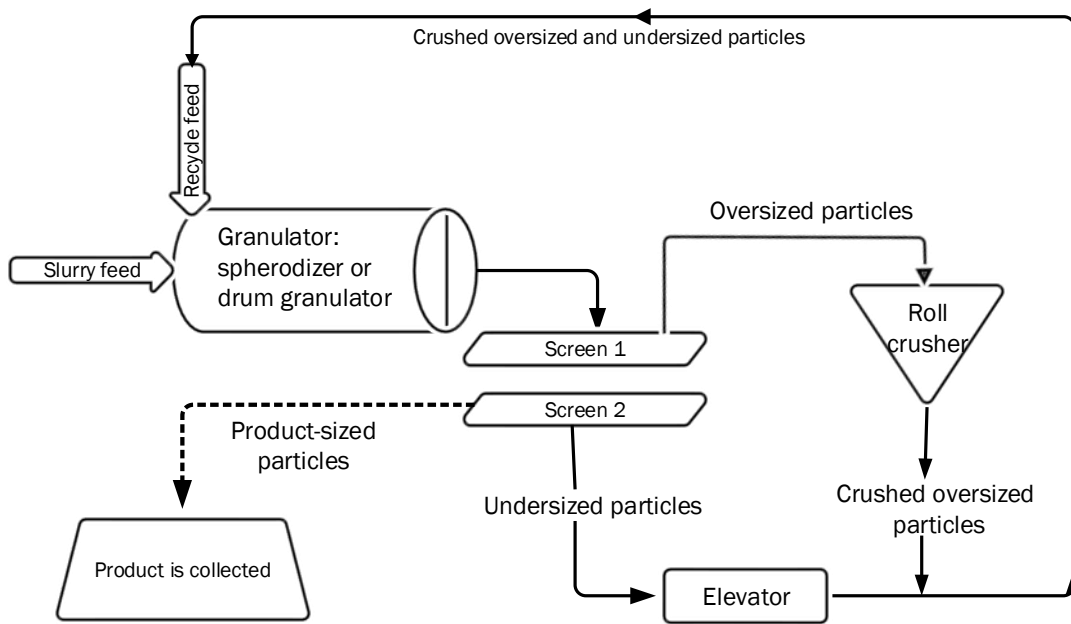
Figure 1. Schematic diagram of granulation loop.
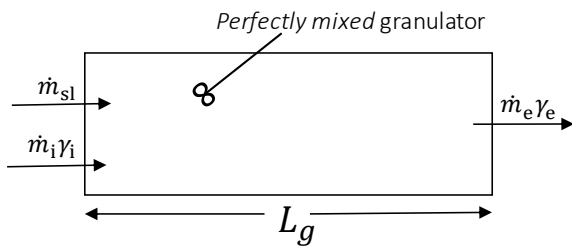
Figure 2. Graphical representation of the *perfectly mixed* granulator.
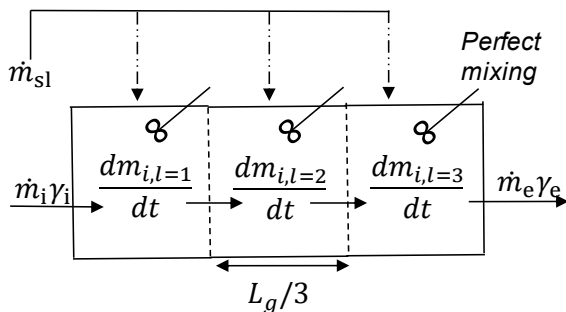
Figure 3. Graphical representation of the multi-compartment granulator. Abreviations: N_z is the number of the compartments in the granulator; L_g is the length of the granulator.

2 Application of PBE to granulation process in spherodizers

In the fertilizer production plant under consideration, a continuous granulation process is used. In an industrial application, it is relatively easier to work with mass based population balance equations (PBEs) instead of number based PBEs due to (i) the PSD in a real plant is typically measured by sieving and weighting, and (ii) mass based PBE is more convenient to use from a numerical point of view due to a huge number of particles compared to their mass. In addition, the size of the particles is represented in terms of their diameters (L) since measuring of PSDs in the plant are based on sieve diameter. A number based PBE with particle volume as internal coordinate is described in (Ramkrishna, 2000). Thus, it is essential to convert PBEs from their volume based formulations to length based formulations. In addition, number based PBEs should be converted into mass based PBEs. The mass based PBE for the spherodizer (continuous layering process) taking particle diameter as the internal coordinate, is formulated as

$$\frac{\partial m(L,z,t)}{\partial t} = -L^3 \frac{\partial}{\partial L}\left[G\frac{m(L,z,t)}{L^3}\right] - \frac{\partial}{\partial z}\left[\frac{dz}{dt}m(L,z,t)\right].$$
$$(1)$$

where m is the mass density function $\left[\frac{kg}{mm^3 \cdot [mm]}\right]$. The first term on the right hand side represents the particle growth due to layering, while the last term represents a continuous process and gives the flow of particles through the granulator. G is the growth rate $\left[\frac{mm}{s}\right]$. Equation 1 is derived by assuming that all particles are ideal spheres
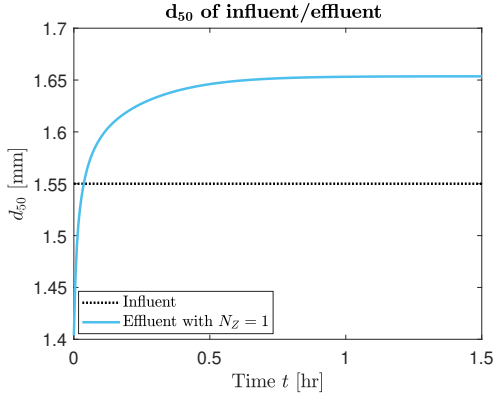
Figure 4. d_{50} for influent and effluent.

with a constant solid density ρ.

Here, the particle growth rate for layering (G) is modeled assuming a linear size-independent growth rate. Mathematical expressions for a linear size-independent growth rate are given in Equation 2.

$$
\begin{aligned}
G &= \frac{2\dot{m}_{\text{sl}}(1 - X_{\text{sl,i}})}{\rho A_{\text{p,tot}}}, \\
A_{\text{p,tot}} &= \pi n \int_{L=0}^{L=\infty} L^2 \, dL,
\end{aligned}
\tag{2}
$$

where $\dot{m}_{\text{sl}}$ is the fresh fertilizer spray rate, $X_{\text{sl,i}}$ is the moisture fraction in the slurry and total $A_{\text{p,tot}}$ is the surface area of the particles.

2.1 Internal coordinate

Integration of Eq. 1 for i-th size class gives

$$
\frac{dM_i}{dt} = L_i^3 G \cdot \left[m\left(t, L_{i-\frac{1}{2}}\right) - m\left(t, L_{i+\frac{1}{2}}\right) \right].
\tag{3}
$$

For simplicity, in Equation 3, the particle flux term is neglected as it is dependent on the external coordinate (the discretization of the external coordinate is described in Section 2.2).

Discretization of the internal coordinate, i.e., discretization of the growth term for particle diameter, has been performed using a finite volume scheme extended by a flux limiter function. High resolution schemes are considered to attain higher accuracy than the first order upwind schemes. In addition, these methods avoid spurious oscillations by applying a high order flux in the smooth regions and a low order flux near discontinuities (Koren, 1993; Kumar, 2006). In this paper, a Koren flux limiter function (KFL) is used to achieve a robust upwind discretization scheme to Eq. 3. Discretization of the internal coordinate is performed on a linear grid using the KFL scheme. KFL scheme for the mass based PBE with the particle diameter (not volume) as internal coordinate is given in (Vesjolaja et al., 2018).

2.2 External coordinate

For a continuous granulation process, a plug flow along the axial direction is assumed. Here, the external coordi-

nate (particle fluxes in and out of the granulator), given by the term $\frac{\partial}{\partial z}\left[\frac{dz}{dt}m(L,z,t)\right]$ is treated in two different ways as described below as Case I and Case II, respectively.

Case I: In this simplified case, a concept of *output equivalent* inside the granulator is used (Figure 2). This means that the entire granulator is treated as *perfectly mixed* throughout its length (L_{g}). The whole granulator is treated as a single compartment with $N_z = 1$ where N_z denotes the number of compartments. Thus, the particle flux term reduces to

$$
\frac{\partial}{\partial z}\left[\frac{dz}{dt}m(L,z,t)\right] = \dot{m}_{\text{i}}\gamma_{\text{i}} - \dot{m}_{\text{e}}\gamma_{\text{e}},
\tag{4}
$$

where the particle flux out from the granulator is defined as

$$
\dot{m}_{\text{e}}\gamma_{\text{e}} = \frac{m_i}{\tau}\gamma_{\text{e}}.
\tag{5}
$$

Here, $\dot{m}_{\text{i}}$ is the mass flow rate of particles entering the granulator (influent), $\dot{m}_{\text{e}}$ is the mass flow rate of particles leaving the granulator (effluent), γ_{i} is the size distribution function of the influent, γ_{e} is the size distribution function of the effluent, m_i is the mass of the i-th particle size class, and τ is the retention (residence) time.

Case II: In this case, the granulator is divided into N_z equally sized compartments, Figure 3. The influent to the granulator enters the 1-st compartment and the effluent leaves the granulator from the N_z-th or the last compartment. The particle flux term is discretized by using one of the finite volume schemes. In this paper, a high resolution scheme with Koren flux limiter function (KFL) is used to discretize the spatial domain. For this, the granulator is divided into N_z uniformly spaced compartments, and each compartment is assumed to be *perfectly mixed*. Integration of the particle flux term for N_z compartments gives

$$
\frac{\partial}{\partial z}\left(\frac{dz}{dt}m_{i,z}\right) = w\frac{\partial}{\partial z}(m_{i,z}) = w\left[m_{i,z-\frac{1}{2}} - m_{i,z+\frac{1}{2}}\right],
\tag{6}
$$

where $\frac{dz}{dt} = w$ is the particle velocity along the granulator and is assumed to be constant inside the granulator. The approximation of the terms $m_{i,z\pm\frac{1}{2}}$ is then performed using a KFL scheme (Koren, 1993). The approximation of the terms $m_{i,z\pm\frac{1}{2}}$ using the KFL scheme is given by Eq. 7 and Eq. 8.

$$
\begin{aligned}
m_{i,z-\frac{1}{2}} \approx \frac{1}{\Delta z}\Bigg\{ &\frac{M_{i,z-1}}{L_i} + \frac{1}{2}\phi\left(\theta_{i-\frac{1}{2}}\right) \\
&\times \left(\frac{M_{i,z-1}}{L_i^3} - \frac{M_{i,z-2}}{L_i^3}\right)\Bigg\},
\end{aligned}
\tag{7}
$$

$$
\begin{aligned}
m_{i,z+\frac{1}{2}} \approx \frac{1}{\Delta z}\Bigg\{ &\frac{M_{i,z}}{L_i} + \frac{1}{2}\phi\left(\theta_{i-\frac{1}{2}}\right) \\
&\times \left(\frac{M_{i,z}}{L_i^3} - \frac{M_{i,z-1}}{L_i^3}\right)\Bigg\},
\end{aligned}
\tag{8}
$$
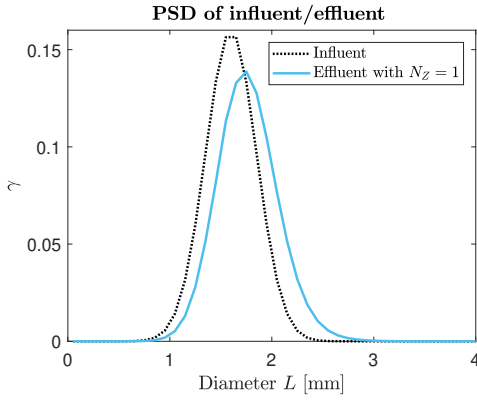
Figure 5. PSD of influent and effluent.

where, M_i is the total mass of the particle in the ith class and $\Delta z = \frac{L_g}{N_z}$ is the length of the each section in the granulator (Figure 3). The limiter function ϕ in Eq. 7 and Eq. 8 is defined as

$$\phi(\theta) = \max\left[0, \min\left(2\theta, \min\left(\frac{1}{3} + \frac{2\theta}{3}, 2\right)\right)\right], \quad (9)$$

and parameter θ is defined as

$$\theta_{i-\frac{1}{2}} = \frac{\frac{M_{i,z}}{L_i^3} - \frac{M_{i,z-1}}{L_i^3} + \varepsilon}{\frac{M_{i,z-1}}{L_i^3} - \frac{M_{i,z-2}}{L_i^3} + \varepsilon}, \quad \theta_{i+\frac{1}{2}} = \frac{\frac{M_{i,z+1}}{L_i^3} - \frac{M_{i,z}}{L_i^3} + \varepsilon}{\frac{M_{i,z}}{L_i^3} - \frac{M_{i,z-1}}{L_i^3} + \varepsilon}. \quad (10)$$

The constant ε is a very small number to avoid division by zero, e.g. $\varepsilon = 10^{-8}$.

3 Application of PBE to granulation process in rotary drums

The model for the rotary drum granulator (continuous agglomeration process) includes binary agglomeration of the particles and is given as

$$\frac{\partial m(L,z,t)}{\partial t} = B(L,z,t) - D(L,z,t) - \frac{\partial}{\partial z}\left[\frac{dz}{dt}m(L,z,t)\right]. \quad (11)$$

To solve the model, the entire particle size range is divided into uniformly distributed classes. The particle flux term (the last term on the right hand side of Eq. 11) is treated in two different ways as described in detail in Section 2.2. The agglomeration terms (the first two terms on the right hand side of Eq. 11) are discretized using the cell average scheme (Kumar et al., 2006; Kumar, 2006) as well as Kumar et al.'s new finite volume scheme (Kumar et al., 2016).

The cell average (CA) scheme was introduced by Kumar (Kumar et al., 2006; Kumar, 2006) and it belongs to the sectional methods of discretization. In the CA scheme, at first the total birth of particles in each cell denoted is computed. Then the average volume of the newly formed

particles in each cell is calculated. The next step in the CA scheme is to assign the total birth of particles appropriately to different cells depending on the position of the average volume of all newborn particles relative to the cell center volume. However, the CA scheme discussed in (Kumar et al., 2006; Kumar, 2006) is valid when the particle volume represents the internal coordinate. Thus, the volume based formulation of the CA scheme should be transformed into the diameter based formulation of the CA scheme. For this, the zeroth moment (total number of particles), and the third moment (total mass of particles) has been chosen to be conserved (compared to zeroth and first moments for volume based definition). Mathematical expressions of diameter based formulation for the CA scheme are given in (Vesjolaja et al., 2018).

Kumar et al.'s new finite volume scheme (Kumar et al., 2016) is based on the finite volume approach proposed by (Forestier and Mancini, 2012). Recently, new approach of solving PBE was proposed in (Kumar et al., 2016). This scheme is an accurate and efficient discretization method for agglomeration tern discretization. It has an improvement over the finite volume scheme proposed by (Forestier and Mancini, 2012) since it provides better solution of several moments in addition to the mass conservation property. Mathematical formulations are given in (Kumar et al., 2016).

Table 1. Parameters used for simulating granulation in spherodizers

Parameter	Spherodizer
Range of L [mm]	0-8
Number of cells	80
Grid type	linear
ρ [kg$\cdot$ m^{-3}]	1300
Length of granulator [m]	10
τ [min]	10
$\dot{m}_{\text{sl,i}}$ [kg$\cdot$ h^{-1}]	100
$X_{\text{sl,i}}$	0.05
Time step for RK4 [s]	20

In this paper, the agglomeration kernel (β) is defined using the Kapur agglomeration kernel model (Kapur, 1972) by taking $a = 2$ and $b = 1$. Using the diameter based formulation, the agglomeration kernel takes the form

$$\beta_{ik} = \left(\frac{6}{\pi}\right)^{\frac{2}{3}} \frac{1}{\rho} \beta_0 K_{ik}, \quad (12)$$

where the term $\left(\frac{6}{\pi}\right)^{\frac{2}{3}} \frac{1}{\rho}$ arises during the conversion from a number-based formulation to the mass-based formulation of PBEs, β_0 is the particle size independent part of the agglomeration kernel, and K_{ik} is the particle size dependent part of the agglomeration kernel as shown in Eq.13.

$$K_{ik} = \frac{(L_i + L_k)^2}{L_i L_k}, \quad (13)$$

PSD at the granulator outlet
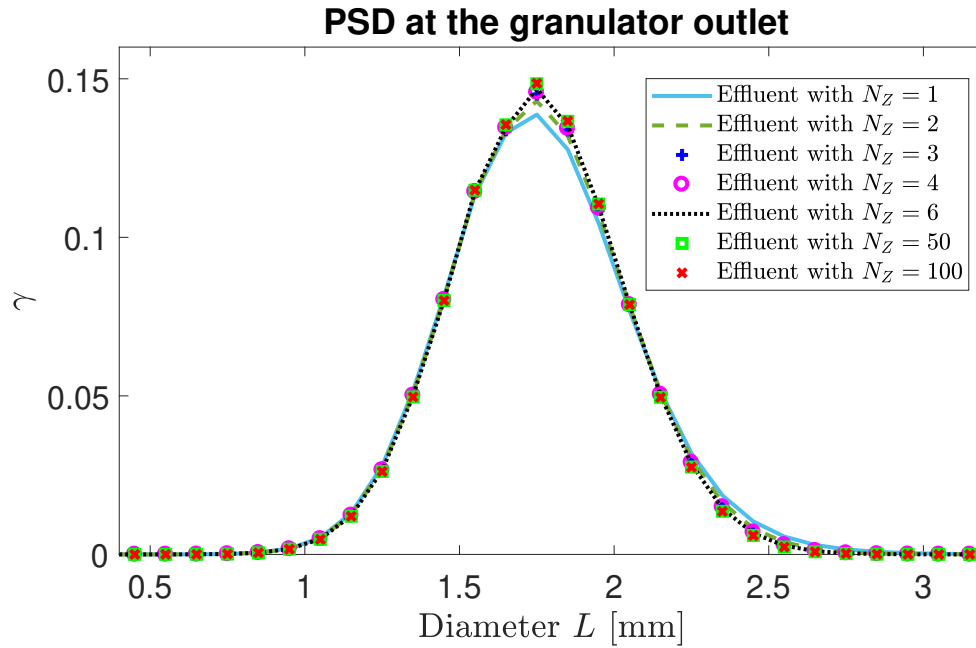
Figure 6. PSDs at the outlet of the spherodizer for different number of compartments.

4 Simulation Results and Discussion

4.1 Simulation Setup

The semi-discrete form (set of ODEs) of the PBEs obtained from particle class size and spatial discretizations are solved using a 4-th order Runge-Kutta method with fixed time step. Dynamic simulations are performed using MATLAB (MATLAB, 2017). The parameters used to simulate the application of PBE to the fertilizer granulation process in spherodizers and rotary drums are given in Table 1 and Table 2 respectively .

Table 2. Parameters used for simulating granulation in rotary drums.

Parameter	Rotary drum
Range of L [mm]	0-8
Number of cells	80
Grid type	linear
ρ [kg· m^{-3}]	1300
β_0 [s^{-1}]	$1.0 \cdot 10^{-11}$
Length of granulator [m]	6
τ [min]	6
$\dot{m}_{\mathrm{sl,i}}$ [kg· h^{-1}]	100
$X_{\mathrm{sl,i}}$	0.05
Time step for RK4 [s]	20

4.2 Simulation results for granulation in spherodizers

The granulation process in spherodizers is simulated for two simulation cases: Case I where the entire granulator

Table 3. Comparison of the computational time (in seconds) with different numerical schemes for rotary drum simulations.

Numerical scheme	$\beta = \beta_0$	$\beta = \beta_{ik}$
CA with $N_z = 1$	3.3	6.0
CA with $N_z = 3$	9.0	17.0
NFV with $N_z = 1$	3.6	6.2
NFV with $N_z = 3$	9.1	17.3

is assumed 'perfectly mixed', and Case II where the entire granulator length is divided into uniformly sized compartments as described in Section 2.2.

The mass based formulation of the PBE with diameter representing the particle size, is used, and the spherodizer is a continuous process with influent and effluent. Table 1 lists the simulation settings and parameters values. Simulation results are presented by particle size distribution (PSD) plots in terms of particle diameter, as well as by particle median diameter d_{50}. The values of d_{50} are obtained from the cumulative mass distribution. Linear interpolation is used to extract d_{50} values that correspond to intercept for 50 % of cumulative mass.

Simulation results obtained for the application of the PBEs in spherodizers is depicted in Figure 4, Figure 5 and Figure 6. Figure 4 and Figure 5 shows the change in particle sizes that occurs during a continuous granulation process in the *perfectly mixed* spherodizer (solution of the PBE is found using KFL scheme). Clearly, the particles grow in size inside the granulator. As a result, d_{50} of the effluent is larger than d_{50} of the influent, Figure 4, and more of coarse-sized particles is produced,
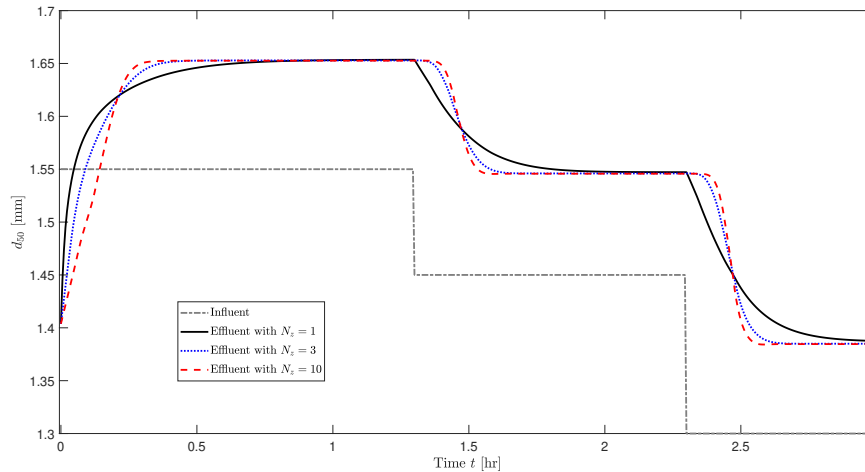
Figure 7. d_{50} at the outlet of the spherodizer for different number of compartments.
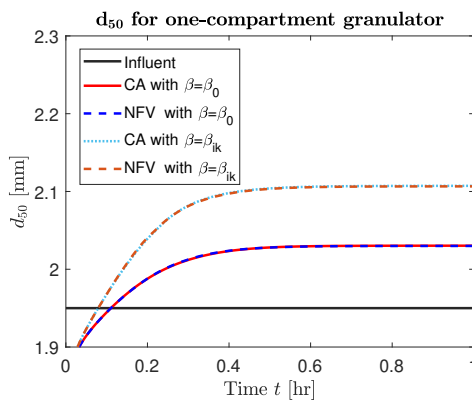
Figure 8. Simulation results for d_{50} with size-dependent and size-independent agglomeration kernels assuming *perfectly mixed* drum granulator.

Figure 5. Simulation results with the inclusion of spatial discretization of the granulator are depicted in Figure 6 and Figure 7. In Figure 6, PSDs of the effluent at the outlet of the granulator with different number of compartments (N_z) are compared. The PSD plot with $N_z = 50$ can be considered as a reference plot since no change in PSD is observed when increasing the number of compartments in the granulator (PSD plot with $N_z = 100$). Differences in PSDs appear mainly for the coarser particle fractions (with $L > 1.5$ mm), while for the finer particles (with $L < 1.5$ mm), the *perfectly mixed* granulator gives accurate enough results. Figure 7 shows d_{50} of the effluent as a step change in the influent is given. In a multi-compartment granulator a time delay is introduced as a step change of d_{50} in influent is given. Thus, inclusion of the spatial discretization could be important for development of a control-relevant dynamic model of the graulator. Based on Figure 6 and Figure 7, it can be concluded that diving the whole granulator space in 3 compartments could be sufficient to achieve a sufficiently accurate model

compared to the *perfectly mixed* granulator ($N_z = 1$). The only disadvantage of discretizing the granulator in space is the increased computational time. As the value of N_z increases, i.e., as the number of compartments in the granulator increases, the computational time also increases. In particular, with a standard PC used for the simulation (i5 with 4 cores, 8 GB RAM and 2.1 GHz CPU), the computational time for one-compartment granulator was 0.5 s, and for the three-compartment granulator was 0.9 s (about 2 times more).

4.3 Simulation results for granulation in rotary drums

The granulation process in drum granulators is simulated using the CA scheme, as well as the NFV scheme. Similar to the spherodizers, the simulations have been performed for a multi-compartment drum granulator model and also with the *perfectly mixed* assumption. The mass based formulation of the PBE with particle diameter representing the size, is used to assess the PSD. The simulation results are also compared for a size-independent ($\beta = \beta_0$) and a size-dependent ($\beta = \beta_{ik}$) agglomeration kernel.

With the *perfectly mixed* assumption, i.e., for $N_z = 1$, the simulation results are depicted in Figure 8. As expected, the CA scheme and the NFV scheme produce similar results (d_{50} of the effluent) for both the size-independent and the size-dependent agglomeration kernels. Due to binary agglomeration, the particles grow in size. However, with the size-dependent agglomeration kernel, the d_{50} of the effluent is higher than with the size-independent constant kernel.

Simulation results for the multi-compartment drum granulator model are shown in Figure 9. Figure 9 compares the PSD of the influent, as well as PSDs of the effluents that correspond to each of the compartments of the granulator. Clearly, particles in the first compartment are smaller in size compared to the second and third compartments of the granulator. Particles increase in their sizes
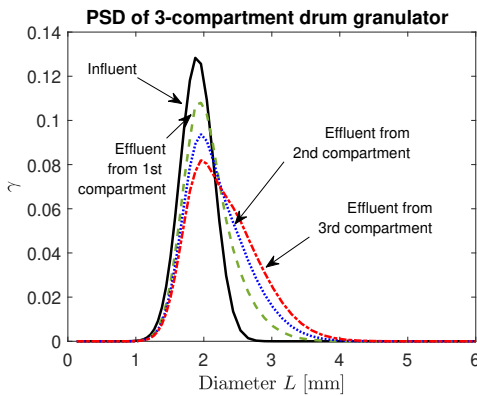
Figure 9. PSD for influent/effluent in 3-compartment drum granulator with the CA scheme.

as they are transported along the granulator, and as a result, the coarser fractions of the particles increase while their finer fractions decrease. The computational time is also compared for different choices of N_z as listed in Table 3. As expected, the computational time needed for solving the model increases with increasing the number of compartments in the drum granulator. This observation is valid with both the CA and the NFV schemes used to solve the model, and for both types of agglomeration kernels. However, inclusion of the size-dependent agglomeration kernel increase the computational time significantly: it takes almost twice the time to obtain the solution with size-dependent agglomeration kernel ($\beta = \beta_{ik}$) compared to the size-independent kernel $\beta = \beta_0$.

5   Conclusions

In this paper, the population balance equations for the spherodizer and the rotary drum granulator were developed. To account for property inhomogeneity in the granulators, multi-compartment models of the granulators were also developed. The simulation results showed that the discretization of the external coordinate (axial length of the granulator) in the model, introduces a transport (time) delay from the inlet of the granulator to its outlet. Inclusion of the correct transport delay is important for control studies. However, the ability of the model to capture transport delay inside the granulator comes with the cost of increased computational time. Two different discretization schemes, namely Kumar's new finite volume scheme and the cell average scheme showed similar simulation results in terms of the model solution accuracy and computational time. Model solution was obtained relatively fast for both simulation scenarios: with the constant agglomeration kernel and with the size-dependent agglomeration kernel. Thus, the developed models and model solution techniques can be used for further control-relevant studies.

6   Acknowledgment

References

I.T. Cameron, F.Y. Wang, C.D. Immanuel, and F. Stepanek. Process systems modelling and applications in granulation: A review. *Chemical Engineering Science*, 60(14):3723–3750, 2005.

L. Forestier and S. Mancini. A finite volume preserving scheme on nonuniform meshes and for multidimensional coalescence. *SIAM Journal of Scientific Computing*, 34(6), 2012. doi:10.1137/110847998.

C. Herce, A. Gil, M. Gil, and C. Cortés. A cape-taguchi combined method to optimize a npk fertilizer plant including population balance modeling of granulation-drying rotary drum reactor. In *Computer Aided Chemical Engineering*, volume 40, pages 49–54. Elsevier, 2017.

PC Kapur. Kinetics of granulation by non-random coalescence mechanism. *Chemical Engineering Science*, 27(10):1863–1869, 1972.

B. Koren. A robust upwind discretization method for advection, diffusion and source terms. In C. B. Vreugdenhil and B. Koren, editors, *Numerical Methods for Advection-Diffusion Problems, Notes on Numerical Fluid Mechanics*, pages 117–138. 1993.

J. Kumar. *Numerical approximations of population balance equations in particulate systems*. PhD thesis, Otto-von-Guericke-Universität Magdeburg, Universitätsbibliothek, 2006.

J. Kumar, M. Peglow, G. Warnecke, S. Heinrich, and L. Mörl. Improved accuracy and convergence of discretized population balance for aggregation: The cell average technique. *Chemical Engineering Science*, 61(10):3327–3342, 2006.

J. Kumar, G. Kaur, and E. Tsotsas. An accurate and efficient discrete formulation of aggregation population balance equation. *Kinetic & Related Models*, 9(2), 2016.

J. Litster and B. Ennis. *The science and engineering of granulation processes*, volume 15. Springer Science & Business Media, 2004.

MATLAB. 2017a. The MathWorks, Inc., Natick, Massachusetts, United States., 2017.

R. Ramachandran, C.D. Immanuel, F. Stepanek, J.D. Litster, and F.J. Doyle III. A mechanistic model for breakage in population balances of granulation: Theoretical kernel development and experimental validation. *Chemical Engineering Research and Design*, 87(4):598–614, 2009.

D. Ramkrishna. *Population balances: Theory and applications to particulate systems in engineering*. Academic press, 2000.

G. Valiulis and R. Simutis. Particle growth modelling and simulation in drum granulator-dryer. *Information Technology and Control*, 38(2), 2009.

L. Vesjolaja, B. Glemmestad, and B. Lie. Population balance modelling for fertilizer granulation process. *Proceedings of The 59th Conference on Simulation and Modelling (SIMS 59), 26-28 September 2018, Oslo Metropolitan University, Norway*, 2018.

F.Y. Wang, X.Y. Ge, N. Balliu, and I.T. Cameron. Optimal control and operation of drum granulation processes. *Chemical Engineering Science*, 61(1):257–267, 2006.

Solving the population balance equation for granulation processes: particle layering and agglomeration

Ludmila Vesjolaja[1]   Bjørn Glemmestad[2]   Bernt Lie[1]

[1]Department of Electrical Engineering, IT and Cybernetics, University of South-Eastern Norway,
`{ludmila.vesjolaja,bernt.lie}@usn.no`
[2]Process Modeling and Control Department, Yara Technology Center, Norway, `bjorn.glemmestad@yara.com`

Abstract

Granulation processes are frequently used in the fertilizer industry to produce different grades of mineral fertilizers. Large recycle ratios and poor product quality control are some of the problems faced by such industries. Thus, for real time model based process control and optimization, it is necessary to find an appropriate numerical scheme can find solution of the model sufficiently accurate and fast. In this study, population balance principles were used to model particle granulation processes. Different numerical schemes were tested to find simple yet sufficiently accurate solution schemes for population balance equation. Numerical schemes were applied to find the solution of both the *layering* term and the *agglomeration* term that appear in the population balance equation. The accuracies of the numerical schemes were assessed by comparing the numerical results with analytical, tractable solutions. Comparison of the accuracy of numerical schemes showed that a high resolution scheme with Koren flux limiter function might be a good choice for the *layering* term discretization, while a cell averaging technique and a new finite volume method of Kumar et al. (2016) produce a sufficiently accurate solution for the *agglomer-ation* term discretization.

Keywords: population balance, numerical scheme, layering, agglomeration

1 Introduction

Granulation is a particle enlargement process during which fine particles and/or atomizable liquids are converted into granules via a series of complex physical processes (Litster and Ennis, 2004). Here, the focus is on modeling a granulation process used for mineral fertilizer production. In the fertilizer industry, depending on desired product properties, different types of granulators are used, e.g., spherodizers and drum granulators.

Formation of the particles (granulation mechanisms) depends on the granulator type and operating conditions. Particle growth due to layering is predominant in spherodizers. Layering is a continuous process (differential growth) during which particle growth occurs due to a successive coating of a liquid phase onto a granule (Litster and Ennis, 2004). In drum granulators, on the other
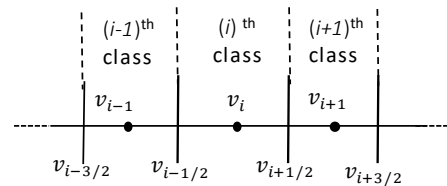
Figure 1. Size discretization into classes (cells) using linear grid.

hand, particle collision occurs, and thus particle agglomeration contributes significantly to particle size change. In this paper, binary particle agglomeration is assumed for population balance (PB) modeling. Binary agglomeration refers to a particle growth mechanism that occurs due to successful collision of two particles, resulting in the formation of a larger, composite particle. Thus, the agglomeration results in a reduction of the total number of particles, while the total mass remains conserved (Litster and Ennis, 2004).

The operation of granulation plants at an industrial scale can be challenging (Litster and Ennis, 2004; Radichkov et al., 2006; Heinrich et al., 2003). Periodic instability associated with the operation of the granulation circuit has been reported. This causes the particle size distribution (PSD) flowing out of the granulator to oscillate, thus making it difficult to maintain the desired product quality, e.g., particle size. Thus, to address and solve these challenges, it is essential to have a dynamic model of the granulator that can further be used to design optimal control structures. The model should be both simple and sufficiently accurate to reflect the underlying physical mechanisms that take place in the granulator. The resulting population balance equations (PBEs) are nonlinear in nature and are challenging to solve. Analytical solutions of these PBEs are available only for ideal and simplified cases, and thus for most of the cases, numerical methods are needed to solve such PBEs. For real time model based process control and optimization, it is necessary to find an appropriate numerical scheme that is sufficiently accurate and fast. Therefore, the main focus

of this paper has been (i) investigation of the accuracy of different numerical schemes that are suitable for discretizing PBEs by comparing the numerical results with analytically tractable solutions; (ii) application of various finite volume techniques (first order upwind, second order central difference, and a high resolution scheme) to the growth by a layering process; (iii) application of different sectional methods (Hounslow method, cell average technique, fixed pivot scheme), and a new finite volume method of Kumar et al. (2016) to the agglomeration process.

2 Population Balance Equation (PBE)

Population balance (PB) is frequently used to describe dynamics of particle property distributions, e.g., particle size distribution, moisture content in particles and porosity (Ramkrishna, 2000). The general form of a number based PBE with particle size (x) as the internal coordinate, spatial variation (z) as the external coordinate, and time t as time coordinate is represented as

$$\frac{\partial n(x,z,t)}{\partial t} = -\frac{\partial}{\partial x}\left[Gn(x,z,t)\right] + B(x,z,t) - D(x,z,t)$$
$$-\frac{\partial}{\partial z}\left[\frac{dz}{dt}n(x,z,t)\right], \quad (1)$$

where $n(x,z,t)$ is the number density function $\left[\frac{no}{mm^3 \cdot [internal\ coordinate]}\right]$. The first term on the right hand side represents the particle growth due to layering, the second and third terms stand for particle birth and death, respectively, due to agglomeration, while the last term represents a continuous process and gives the flow of particles through the granulator. G is the growth rate $\left[\frac{internal\ coordinate}{s}\right]$ (Ramkrishna, 2000). The birth B and death D terms usually include integrals that lead to partial integro-differential equations which make the solution of the PBEs complicated. Mathematical expressions for birth and death terms are shown in Section 4 when describing different numerical schemes for binary agglomeration. Further simplifications of the general PBE (Eq. 1) are possible and are dependent on the nature of the process taken into consideration.

At this point, it is convenient to define moments of particle size distributions that will be used later for discussion of the simulation results. The l-th moment of the PSD is defined as

$$\mu^l = \int_0^\infty x^l n(x)dx. \quad (2)$$

The first moments are of particular interest. Depending on the choice of internal coordinate (e.g., particle volume or particle length) the moments are related to the total number, length, area, and volume of particles.

3 Numerical schemes for *layering* term discretization

PBEs are non-linear in general and analytical solutions are available only for simplified processes. Thus, a simple 1-D batch process for which an analytical solution is available is modeled so that the accuracy of the numerical schemes can be reliably evaluated. In a batch granulation process there is no continuous particle flow through the granulator. If the particle size change in the granulator is mainly due to layering (e.g., in spherodizers), and the size of a particle is represented by its volume v, then Eq. 1 reduces to

$$\frac{\partial n(v,t)}{\partial t} = -\frac{\partial}{\partial v}\left[Gn(v,t)\right]. \quad (3)$$

In Eq. 3, the concept of *perfect mixing* inside the granulator is applied: particle property (e.g., size distribution) inside the granulator is the same at every point inside the granulator. In this paper, the solution to PBEs containing growth term G (Eq. 3) is found by transforming the partial differential equation (PDE) into a system of ordinary differential equations (ODEs), i.e., by reducing the dimensionality of the problem with respect to the particle size. The set of ODEs can then be solved using an appropriate time integrator. In this paper, a Runge-Kutta 4-th order (RK-4) time integration method is used for all simulations.

For particle size discretization, first the particles are classified into N_c particle classes which are numbered by $i \in \{1,2,...N_c\}$ classes (cells) using a *linear grid* as shown in Figure 1.

Here, i represents the i-th particle class, v_i is the volume of the particle of the i-th class, $v_{i\pm\frac{1}{2}}$ is the left and the right boundaries of the i-th class, and $\triangle v = v_{i+\frac{1}{2}} - v_{i-\frac{1}{2}}$ is the size of the classes. The dots in each class (Figure 1) represent the cell centers. Secondly, an appropriate numerical scheme is applied to convert Eq. 3 into set of ODEs. Integration of Eq. 3 over cell i from $v_{i-\frac{1}{2}}$ to $v_{i+\frac{1}{2}}$ gives

$$\frac{dN_i(t)}{dt} = G\left(v_{i-\frac{1}{2}}\right)n\left(v_{i-\frac{1}{2}},t\right) - G\left(v_{i+\frac{1}{2}}\right)n\left(v_{i+\frac{1}{2}},t\right). \quad (4)$$

In a simplified case, a solution of Eq. 4 can be found analytically: The particle growth due to layering in a batch process does not change the total number of particles in the batch, but only the particle volume is changed. If t denotes the time for particle growth, v_{initial} denotes the initial volume of the particles (for all classes), and G is the constant growth rate, then the new volume (v_{new}) of the particles after the growth due to layering is given as

$$v_{\text{new}} = v_{\text{initial}} + t \cdot G. \quad (5)$$

However, in real applications, analytical solutions are difficult to obtain, and, thus various numerical schemes are applied to approximate the right hand side of Eq. 4. The

PBE represented by Eq. 3 is a hyperbolic equation due to the layering term, and Eq. 3 can be approximated using a finite volume scheme that automatically incorporates conservation of number in a growth process.

In this paper, three finite volume schemes are compared for particle size discretization, namely a first order upwind scheme (FU) and a second order central difference scheme (SCD). The first order upwind (FU) scheme uses the approximation defined by Eq. 6 and 7:

$$n\left(v_{i-\frac{1}{2}},t\right) \approx \frac{1}{\Delta v}\left[N_{i-1}(t)\right], \qquad (6)$$

and

$$n\left(v_{i+\frac{1}{2}},t\right) \approx \frac{1}{\Delta v}\left[N_i(t)\right]. \qquad (7)$$

Thus, approximation of Eq. 4 using the FU scheme leads to

$$\frac{dN_i}{dt} \approx \frac{1}{\Delta v}\left[G\left(v_{i-\frac{1}{2}}\right)N_{i-1}(t) - G\left(v_{i+\frac{1}{2}}\right)N_i(t)\right]. \qquad (8)$$

Assuming constant growth rate in all cells, Eq. 8 simplifies to

$$\frac{dN_i^{\text{FU}}}{dt} \approx \frac{G}{\Delta v}\left[N_{i-1}(t) - N_i(t)\right]. \qquad (9)$$

The second order central difference (SCD) scheme uses the approximation

$$n\left(v_{i-\frac{1}{2}},t\right) \approx \frac{1}{\Delta v}\frac{[N_{i-1}(t) + N_i(t)]}{2}, \qquad (10)$$

and

$$n\left(v_{i+\frac{1}{2}},t\right) \approx \frac{1}{\Delta v}\frac{[N_i(t) + N_{i+1}(t)]}{2}. \qquad (11)$$

Thus, discretization of Eq. 4 over a cell i using the SCD scheme results in

$$\frac{dN_i^{\text{SCD}}}{dt} \approx \frac{G}{\Delta v}\frac{[N_{i-1}(t) - N_{i+1}(t)]}{2}, \qquad (12)$$

where particle growth G is assumed constant in all the cells.

A finite volume scheme that is extended by a flux limiter is also applied to Eq. 4 to reduce the dimensionality of the PBEs with respect to the particle size. In particular, the Koren flux limiter function (Koren, 1993) is used to achieve a robust upwind discretization scheme to Eq. 4. High resolution schemes are considered to attain higher accuracy than the first order upwind schemes. In addition, these methods avoid spurious oscillations by applying a high order flux in the smooth regions and a low order flux near discontinuities (Koren, 1993; Kumar, 2006). Assuming constant G in all cells, Eq. 4 can be discretized with the Koren scheme as

$$\frac{dN_i^{\text{KFL}}}{dt} \approx G \cdot \left[n\left(v_{i-\frac{1}{2}}\right) - n\left(v_{i+\frac{1}{2}}\right)\right], \qquad (13)$$

where,

$$n\left(v_{i-\frac{1}{2}},t\right) \approx \frac{1}{\Delta v}[N_{i-1}(t) \\ \times \frac{1}{2}\phi\left(\widetilde{\theta}_{i-\frac{1}{2}}\right) \cdot (N_{i-1}(t) - N_{i-2}(t))], \qquad (14)$$

and

$$n\left(v_{i+\frac{1}{2}},t\right) \approx \frac{1}{\Delta v}\left[N_i(t) + \frac{1}{2}\phi\left(\widetilde{\theta}_{i+\frac{1}{2}}\right) \cdot (N_i(t) - N_{i-1}(t))\right]. \qquad (15)$$

Here, ϕ is the limiter function defined as

$$\phi\left(\widetilde{\theta}\right) = \max\left[0, \min\left(2\widetilde{\theta}, \min\left(\frac{1}{3} + \frac{2\widetilde{\theta}}{3}, 2\right)\right)\right]. \qquad (16)$$

Parameter $\widetilde{\theta}$ is defined as,

$$\widetilde{\theta}_{i-\frac{1}{2}} = \frac{N_i - N_{i-1} + \chi}{N_{i-1} - N_{i-2} + \chi}, \quad \widetilde{\theta}_{i+\frac{1}{2}} = \frac{N_{i+1} - N_i + \chi}{N_i - N_{i-1} + \chi}, \qquad (17)$$

with a very small constant χ (e.g., 10^{-8}) to avoid division by zero.

4 Numerical schemes for *agglomeration* term discretization

PBE for a batch agglomeration process using the particle volume as internal coordinate is given by

$$\frac{\partial n(v,z,t)}{\partial t} = B(v,z,t) - D(v,z,t). \qquad (18)$$

Here, the particle birth (B) and death (D) due to binary agglomeration are modeled using the Hulburt and Katz formulation (Hulburt and Katz, 1964). For a pure agglomeration process, the Hulburt and Katz equation (Hulburt and Katz, 1964) is given as

$$\frac{\partial n(t,v)}{\partial t} = B - D = \\ \frac{1}{2}\int_0^v \beta(t, v-\varepsilon, \varepsilon) n(t, v-\varepsilon) n(t,\varepsilon) d\varepsilon \\ - n(t,v)\int_0^\infty \beta(t,v,\varepsilon) n(t,\varepsilon) d\varepsilon. \qquad (19)$$

Equation 19 represents a 1-D batch process assuming *perfect mixing* inside the granulator. In Eq. 19, β is the agglomeration rate (kernel) that defines the collision frequency of the two particles with volumes v and $v - \varepsilon$.

Agglomeration is a discrete event, and PB modeling of the agglomeration process results in partial integro-differential equations. The integral function appears in the birth (B) and death (D) terms in Eq. 19. Such systems are difficult to solve, and analytical solutions are available only for a limited number of simplified problems. Some of the analytical solutions for different initial conditions and different agglomeration kernels (e.g., constant,

sum and product kernels) are given in Scott (Scott, 1968). Here, for simplicity, the performance of different numerical schemes has been assessed using a constant agglomeration kernel with an exponential initial distribution:

$$n(v,0) = \frac{N_0}{v_0} \exp\left(\frac{-v}{v_0}\right), \qquad (20)$$

is given as

$$n(v,t) = \frac{4N_0}{v_0(\varpi+2)^2} \exp\left(\frac{-2\kappa}{(\varpi+2)}\right), \qquad (21)$$

In Equations 20 and 21, N_0 and v_0 represent the initial number of particles per unit volume and initial mean volume of the particles. The dimensionless volume unit κ, and the dimensionless time variable ϖ are given by Eq. 22:

$$\kappa = \frac{v}{v_0}, \quad \text{and } \varpi = N_0 \beta_0 t. \qquad (22)$$

As a result of particle agglomeration, the total number of particles reduces while the total mass remains constant. The main challenge is to find/develop approximation techniques that would assign the newborn particles accurately while conserving the chosen moments. To achieve this, various numerical methods for expressing the agglomeration term in PBEs are developed, among others the method of moments, the method of successive approximations, the finite volume methods, and the sectional methods. A review of various numerical techniques is summarized in (Ramkrishna, 2000). This paper focuses on applying different sectional methods, as well as a newly developed finite volume technique by Kumar et al.'s (Kaur et al., 2017; Singh et al., 2015; Kumar et al., 2016).

Approximation of the continuous size distribution by a finite number of size sections (cells) has been made using a *geometric type grids*, i.e., the whole particle size interval is divided into a finite number of cells (classes) using geometric progression. This type of grid has been chosen because Hounslow's discretization method, one of the numerical schemes being compared, can be applied only to geometric type grids. The choice of numerical schemes has been made based on suitability for further application of the model for control purposes, i.e., the scheme should posses simplicity in implementation and be sufficiently fast (low computation time), while producing relatively accurate numerical results.

4.1 Hounslow's scheme

According to Hounslow's discretization scheme (Hounslow et al., 1988), the approximation of the continuous size distribution by a finite number of cells is performed using a geometric grid with a factor of two in size, i.e., $v_{i+1} = 2v_i$. Hounslow's discretization scheme (H) is based on four binary interaction mechanisms that can contribute to changes of particles number in the i-th cell. Two of these four mechanisms change the number of particles due

to particle births in the i-th cell, while the other two mechanisms contribute for particle deaths in the i-th cell. Application of the H scheme to the agglomeration process (Eq. 19) gives the total rate of change of particles in each i-th cell as

$$\frac{dN_i^H}{dt} = \sum_{j=1}^{i-2} 2^{j-i+1} \beta_{i-1,j} N_{i-1} N_j + \frac{1}{2} \beta_{i-1,i-1} N_{i-1}^2 -$$
$$N_i \sum_{j=1}^{i-1} 2^{j-i} \beta_{i,j} N_j - N_i \sum_{j=i}^{N_c} \beta_{i,j} N_j. \qquad (23)$$

Here, the first term on the right hand side represents the births of particles that are formed due to collision of particles in the $(i-1)$-th cell with the particles from the first to the $(i-2)$-th cells. The second term on the right hand side stands for the births of particles that are born in cell i by the collision between two particles in the $(i-1)$-th cell. The last two terms in Eq. 23 accounts for the death of particles in the i-th cell.

4.2 Cell average scheme

The cell average (CA) scheme was introduced by Kumar (Kumar et al., 2006; Kumar, 2006). The CA scheme can be applied to both a geometric grid and a linear grid. Here, a geometric grid discretization has been chosen to be able to compare simulation results with the Hounslow's scheme. In the CA scheme, at first the total birth of particles in each cell denoted by B_i is computed:

$$B_i = \sum_{j=1}^{N_{c,i}} B_i^j = \sum_{j,k}^{j \geq k} \left(1 - \frac{1}{2}\delta_{jk}\right) \beta_{jk} N_j N_k, \qquad (24)$$

where the two aggregating particles with volumes v_j and v_k should fulfill the condition $v_{i-\frac{1}{2}} \leq v_j + v_k \leq v_{i+\frac{1}{2}}$; δ_{jk} is the delta Dirac function such that $\delta_{jk} = 1$ for $j = k$, otherwise $\delta_{jk} = 0$; β_{jk} is the agglomeration kernel for binary agglomeration of particles from the j-th and the k-th cells.

Then the average volume of the newly formed particles in each cell denoted by $\bar{v}_i$ is calculated. The average volume of the particle is then given as

$$\bar{v}_i = \frac{\sum_{j=1}^{N_{c,i}} v_i^j B_i^j}{B_i} = \left[\frac{\sum_{j,k}^{j \geq k} \left(1 - \frac{1}{2}\delta_{jk}\right) \beta_{jk} N_j N_k (v_j + v_k)}{\sum_{j,k}^{j \geq k} \left(1 - \frac{1}{2}\delta_{jk}\right) \beta_{jk} N_j N_k}\right]. \qquad (25)$$

The next step in the CA scheme is to assign the total birth of particles B_i appropriately to different cells depending on the position of the average volume of all newborn particles relative to the cell center volume v_i. In total, there are four birth contributions at node v_i: two coming from the i-th cell itself (when $\bar{v}_i < v_i$ and $\bar{v}_i > v_i$), and two from the neighboring $(i-1)$-th and $(i+1)$-th cells (when $\bar{v}_{i-1} > v_{i-1}$ and $\bar{v}_{i+1} < v_{i+1}$). To combine all the possible birth contributions, for convenience the dimensionless term $\lambda_i^\pm(v)$ is introduced:

$$\lambda_i^\pm(v) = \frac{v - v_{i\pm1}}{v_i - v_{i\pm1}}. \qquad (26)$$

The discretized PBE using the CA technique takes the form of

$$\frac{dN_i}{dt} = B_i^{\text{CA}} - D_i^{\text{CA}}, \quad (27)$$

where

$$
\begin{aligned}
B_i^{\text{CA}} &= B_{i-1}\lambda_i^-\left(\bar{v}_{i-1}\right)H\left(\bar{v}_{i-1}-v_{i-1}\right) \\
&\quad + B_i\lambda_i^-\left(\bar{v}_i\right)H\left(v_i-\bar{v}_i\right) + B_i\lambda_i^+\left(\bar{v}_i\right)H\left(\bar{v}_i-v_i\right) \\
&\quad + B_{i+1}\lambda_i^+\left(\bar{v}_{i+1}\right)H\left(v_{i+1}-\bar{v}_{i+1}\right),
\end{aligned}
\quad (28)
$$

and

$$D_i^{\text{CA}} = N_i \sum_{k=1}^{N_c} \beta_{i,k}N_k. \quad (29)$$

Here, the Heaviside step function H is given as

$$
H(v) = \begin{cases} 1, & \text{if } v > 0 \\ \frac{1}{2}, & \text{if } v = 0 \\ 0, & \text{if } v < 0. \end{cases}
\quad (30)
$$

4.3 Fixed pivot scheme

The fixed pivot (FP) technique was developed by Kumar and Ramkrishna (Kumar and Ramkrishna, 1996). The scheme is based on the Hounslow's method, however, the disadvantage of using only geometric grids with Hounslow's method is eliminated in the fixed pivot scheme. This scheme can be used with any type of grid including linear grids. Here, for comparison of numerical solutions, the same geometric grid as in the Hounslow and CA schemes has been chosen. The main difference between the FP and CA schemes is in assigning the new-born particles to the cells. In the FP scheme, each individual birth in a cell is directly assigned to the appropriate cells, unlike in the CA scheme where the average volume of all the newly born particles is first calculated, and then only the assignment of the particles is performed.

The discrete form of the PBE with the FP scheme is then written as (Kumar and Ramkrishna, 1996)

$$\frac{dN_i}{dt} = B_i^{\text{FP}} - D_i^{\text{FP}}, \quad (31)$$

where

$$B_i^{\text{FP}} = \sum_{j,k}^{j \geqslant k}\left(1-\frac{1}{2}\delta_{jk}\right)\eta\left(\tilde{v}\right)\beta_{jk}N_jN_k, \quad (32)$$

such that $v_{i-1} \leq v_j + v_k \leq v_{i+1}$ and

$$D_i^{\text{FP}} = D_i^{\text{CA}} = N_i \sum_{k=1}^{N_c} \beta_{i,k}N_k. \quad (33)$$

Here, $\tilde{v} = v_k + v_j$ and $\beta_{j,k} = \beta(t, v_j, v_k)$ is the agglomeration kernel. The expression $\eta(\tilde{v})$ for each particle birth assignment is given by Eq. 34,

$$
\eta(\tilde{v}) = \begin{cases} \frac{v_{i+1}-\tilde{v}}{v_{i+1}-v_i}, & v_i \leq \tilde{v} < v_{i+1} \\ \frac{\tilde{v}-v_{i-1}}{v_i-v_{i-1}}, & v_{i-1} \leq \tilde{v} < v_i. \end{cases}
\quad (34)
$$
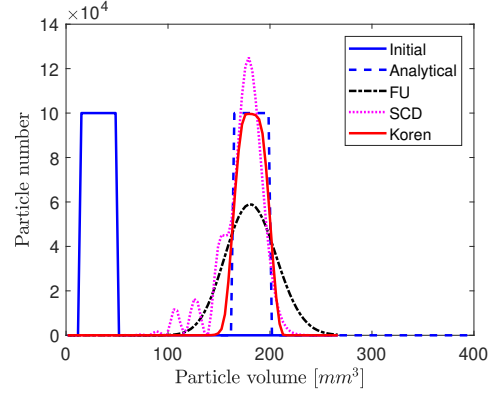
Figure 2. Comparison of PSDs with different numerical schemes for a layering batch process.

4.4 Kumar et al.'s new finite volume scheme

Recently, an accurate and efficient discretization method for agglomeration PBE was proposed in (Kumar et al., 2016) based on the finite volume approach. This scheme has an improvement over the finite volume scheme proposed by (Forestier and Mancini, 2012). The scheme provides better solution of several moments in addition to the mass conservation property compared to the scheme in (Forestier and Mancini, 2012). The Kumar et al.'s new finite volume (NFV) scheme assumes the number density function as the point masses concentrated on the cell representatives. The discrete PBE using Kumar et al.'s (Ku-

Table 1. Parameters used to solve PBE for particle layering process.

Parameter	Layering
Range of v [mm^3]	0-400
Number of cells	80
Grid type	linear
G [mm$^3 \cdot$ s^{-1}]	1
Time step for RK4 [s]	0.1

mar et al., 2016) scheme is then given as

$$\frac{dn_i^{\text{NFV}}}{dt} = \frac{1}{2}\sum_{(j,k)\in Q^i}\beta_{j,k}n_jn_k\frac{\Delta v_j\Delta v_k}{\Delta v_i}S_{i,j,k} - \sum_{j=1}^{N_c}\beta_{i,j}n_in_j\Delta v_j, \quad (35)$$

where factor $S_{i,j,k}$ accounts for mass conservation and is defined as

$$S_{i,j,k} = \frac{v_j + v_k}{v_i}. \quad (36)$$

The set Q^i be a set that contains the pair of cells j and k such that the sum of the cell's representatives, $v_j + v_k$, falls in the domain of cell i represented by cell node v_i:

$$Q^i = \{(j,k) \in N_c \times N_c : v_{i-\frac{1}{2}} < v_j + v_k \leq v_{i+\frac{1}{2}}\}. \quad (37)$$

In the NFV scheme, to ensure that no mass leaves the
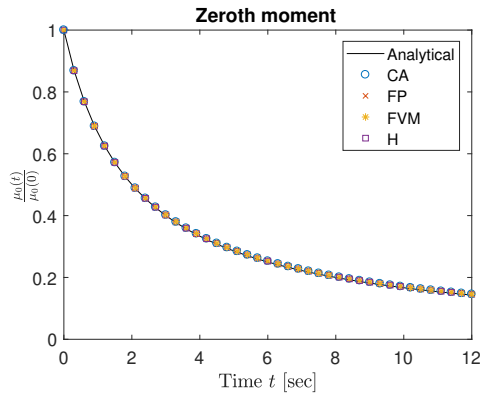
Figure 3. Numerical results for zeroth moment using various numerical schemes for particle agglomeration.
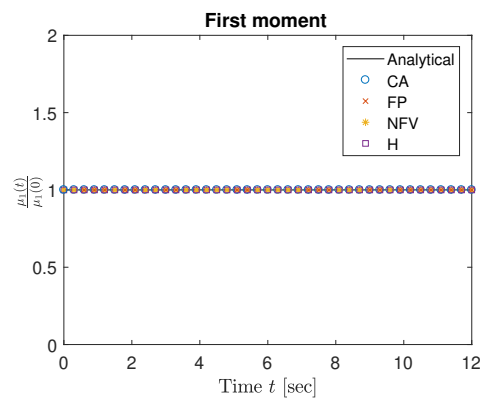
Figure 4. Mass conservation with various numerical schemes.

upper boundary of the size domain (for mass conservation),

$$\beta(v_j, v_k) = \begin{cases} \beta(v_j, v_k), & (v_j + v_k) \le v_{\max} \\ 0, & \text{otherwise}, \end{cases} \quad (38)$$

where $v_{\max}$ is the maximum particle volume.

Table 2. Parameters used to solve PBE for binary particle agglomeration process.

Parameter	Agglomeration
Range of v [mm^3]	0-400
Number of cells	15
Grid type	geometric
β_0 [s^{-1}]	1
Time step for RK4 [s]	0.1
N_0	1
v_0 [mm^3]	1

5 Simulation Results and Discussion

5.1 Simulation Setup

The comparison of different numerical schemes is performed by applying corresponding discretization methods
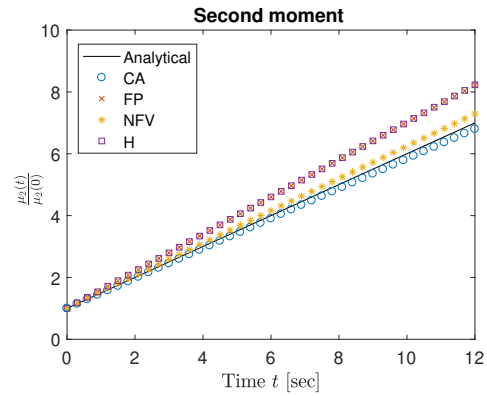
Figure 5. Numerical results for second moment using the various numerical schemes for pure agglomeration.

to batch processes as described in Section 3 and Section 4. Only batch processes with number based PBE with volume as the particle size, are used in these simulations. Such simplified processes are used in order to asses the performance of numerical schemes by comparing their numerical solutions with the analytically tractable solution.

The semi-discrete form (set of ODEs) of the PBEs obtained from particle class size are solved using a 4-th order Runge-Kutta method with fixed time step. Dynamic simulations are performed using MATLAB (MATLAB, 2017). Chosen simulation settings used to asses the accuracy of the numerical schemes for the particle layering process are given in Table 1 while for the particle agglomeration process in Table 2.

5.2 Comparison of numerical solutions for layering process

In order to compare the numerical schemes for particle growth due to layering, the number based PBE is utilized as discussed in Section 3. All the simulation cases for the granulation process due to layering are carried out using linear grid for particle size discretization. The entire particle size range is divided into 80 uniformly distributed cells (classes) and a constant growth rate is assumed (see Table 1). The initial PSD distribution is chosen as

$$N(v, 0) = \begin{cases} 10 \times 10^4 & 15 \le v \le 50, \\ 0 & \text{otherwise}. \end{cases} \quad (39)$$

Three finite volume schemes, i.e., first order upwind (FU), second order central difference (SCD), and Koren flux limiter (KFL) schemes are applied to simulate a pure layering batch process. In addition, simulations with an analytically tractable solution (Eq. 5) are carried out to reliably evaluate the accuracy of the numerical schemes. Simulation results (Figure 2) showed that the resulting particle number distribution obtained using the FU scheme has a diffusive behavior but the solution is stable and smooth. Since a significantly smeared solution is observed, the accuracy of the FU scheme is considered to be low. The numerical result obtained with the SCD scheme

(Figure 2) should have a higher accuracy compared to the FU scheme owing to its higher order. This, however, is true only for smooth solutions. The SCD scheme produces oscillations when the solution changes abruptly, i.e., in the presence of discontinuous analytical solutions. Fortunately, the oscillations that appear in the solution with the SCD scheme do not appear in the numerical solution when the KFL scheme was used. Thus, the highest accuracy among the three tested finite volume schemes is achieved using the finite volume scheme extended with the Koren flux limiter, Figure 2. Higher accuracy of the latter scheme is achieved since in the smooth solution regions, the Koren flux limiter function (in general a second order flux limiter) shows second order accuracy, while in the region where discontinuities appear, the flux limiter acts like a first order scheme, and thus produces a smooth solution. The drawback of the solution obtained with the KFL scheme is the increased computational time: the FU and SCD schemes can be solved relatively faster than the KFL scheme.

5.3 Comparison of numerical solutions for agglomeration process

Comparison of numerical schemes for the pure agglomeration process is carried out on a geometric grid with a factor of 2 in size ($v_{i+1} = 2v_i$) using 15 cells (Table 2).

The four different numerical schemes that are applied to the process for comparison are: Hounslow (H), fixed pivot (FP), cell average (CA), and a new Kumar et al.'s (2016) finite volume (NFV) scheme. The moments are calculated using discrete form of Eq. 2:

$$\mu^l = \sum_i^{Nc} x^l \Delta x_i n_i \tag{40}$$

Assessment of the accuracy of the numerical schemes is performed by comparing the numerical solutions with an analytically tractable solution as mentioned in Section 4.

Simulation results showed that three tested numerical schemes produce solutions where conservation of the zeroth moment (total number conservation) and the first moment (mass conservation) are fulfilled (Figure 3 and Figure 4). No deviation is observed between the analytical solution and the numerical solutions for the first two moments for any of the numerical schemes.

However, the discrepancy between the numerical solutions and the analytical solution shows up for higher moments (second moment, as shown in Figure 5). The best accuracy for the second moment is observed for the numerical solution using the CA scheme.

The NFV scheme also predicts the second moment sufficiently accurate. The poorest solution accuracy for the second moment is produced when the H and the FP schemes are used. In addition, these two sectional schemes produce the same simulation results. This is due to the specific choice of the grid for particle size discretization: The FP scheme becomes equivalent to the H
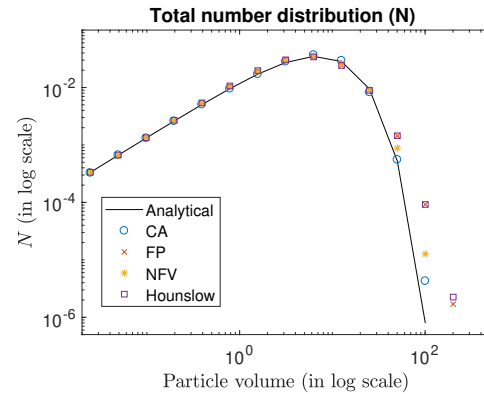
Figure 6. Simulation results for the total number distribution.

scheme when the geometric grid is used with a factor of 2 (Kumar, 2006). Thus, the FP and H schemes should produce the same results when such geometric discretization is used.

Another significant task is to evaluate the performance of the numerical schemes to predict the particle number distribution, usually visualized in log scale as shown in Figure 6. As the particles flow towards the higher size range in the agglomeration process, it is interesting to see how the numerical schemes perform at higher particle size range (thus log scale is used for improved visualization). Here, again, we can see the the discrepancy between the numerical and the analytical solutions. For lower particle size range, all the numerical schemes predicted accurately the particle number distribution. However, for higher particle size range, deviations from the analytical solution are observed. Among the tested numerical schemes, the FP and H schemes over-predict the actual results and thus show the poorest prediction of the particle number distribution. Similar to the moments prediction, the CA scheme and the NFV scheme show relatively better agreement with the analytical solution even for a coarse grid (15 cells).

6 Conclusions

Comparison of numerical schemes for solving population balance equations is presented in this paper, for pure layering and pure agglomeration problems. Discretization of the *layering* term is performed by applying various finite volume schemes. Among the three tested approximation schemes, the Koren flux limiter scheme exhibits relatively better performance in terms of accuracy. However, the Koren scheme also needs a higher computational time compared to the other tested numerical schemes. Numerical solutions for the agglomeration process were obtained by applying different sectional methods, as well as a recent finite volume scheme. The numerical performance of the cell average scheme and the new finite volume scheme in predicting the particle size distribution at higher particle size range is relatively better than Hounslow's scheme and the fixed pivot scheme. The former schemes produce a

PSD that is in good agreement with the analytical solution with coarser grid.

7 Acknowledgment

References

L. Forestier and S. Mancini. A finite volume preserving scheme on nonuniform meshes and for multidimensional coalescence. *SIAM Journal of Scientific Computing*, 34(6), 2012. doi:10.1137/110847998.

S. Heinrich, M. Peglow, M. Ihlow, and L. Mörl. Particle population modeling in fluidized bed-spary granulaiton - analysis of the steady state and unsteady behavior. *Powder Technology*, 130:154–161, 2003. doi:10.1016/S0032-5910(02)00259-0.

MJ Hounslow, RL Ryall, and VR Marshall. A discretized population balance for nucleation, growth, and aggregation. *AIChE Journal*, 34(11):1821–1832, 1988.

H.M. Hulburt and S. Katz. Some problems in particle technology: A statistical mechanical formulation. *Chemical Engineering Science*, 19(8):555–574, 1964.

G Kaur, J. Kumar, and S. Heinrich. A weighted finite volume scheme for multivariate aggregation population balance equation. *Computers & Chemical Engineering*, 101:1–10, 2017.

B. Koren. A robust upwind discretization method for advection, diffusion and source terms. In C. B. Vreugdenhil and B. Koren, editors, *Numerical Methods for Advection-Diffusion Problems, Notes on Numerical Fluid Mechanics*, pages 117–138. 1993.

J. Kumar. *Numerical approximations of population balance equations in particulate systems*. PhD thesis, Otto-von-Guericke-Universität Magdeburg, Universitätsbibliothek, 2006.

J. Kumar, M. Peglow, G. Warnecke, S. Heinrich, and L. Mörl. Improved accuracy and convergence of discretized population balance for aggregation: The cell average technique. *Chemical Engineering Science*, 61(10):3327–3342, 2006.

J. Kumar, G. Kaur, and E. Tsotsas. An accurate and efficient discrete formulation of aggregation population balance equation. *Kinetic & Related Models*, 9(2), 2016.

S. Kumar and D. Ramkrishna. On the solution of population balance equations by discretization - i. a fixed pivot technique. *Chemical Engineering Science*, 51(8):1311–1332, 1996.

J. Litster and B. Ennis. *The science and engineering of granulation processes*, volume 15. Springer Science & Business Media, 2004.

MATLAB. 2017a. The MathWorks, Inc., Natick, Massachusetts, United States., 2017.

R. Radichkov, T. Müller, A. Kienle, S. Heinrich, M. Peglow, and L. Mörl. A numerical bifurcation analysis of continuous fluidized bed spray granulator with external product classification. *Chemical Engineering and Processing*, 45:826–837, 2006. doi:10.1016/j.cep.2006.02.003.

D. Ramkrishna. *Population balances: Theory and applications to particulate systems in engineering*. Academic press, 2000.

WT Scott. Analytic studies of cloud droplet coalescence i. *Journal of the atmospheric sciences*, 25(1):54–65, 1968.

M. Singh, J. Kumar, and A. Bück. A volume conserving discrete formulation of aggregation population balance equations on non-uniform meshes. *IFAC-PapersOnLine*, 48(1):192–197, 2015.

Fertilizer Granulation: Comparison of Modeling Languages

Amir Farzin, Ludmila Vesjolaja, and Bernt Lie

University of South-Eastern Norway, Porsgrunn, Norway, `Bernt.Lie@usn.no`

Abstract

Industrially produced fertilizers are of key importance to produce enough food for a growing global population. On-going work deals with models of the granulation loop in fertilizer production, based on a population balance that finds the particle size distribution of the product. The model is intended for control design in order to dampen or remove production oscillation for reduced energy consumption and improved product quality. In this paper, efficiency of model implementation is studied in addition to the possibility to automate the computation of a linear approximation of the model for control synthesis.

In the implementation study, the current tailor-made MATLAB solver for the model was cloned in computer language Julia. In addition, the implementations in both languages (MATLAB, Julia) were rewritten in a form that allows for use of the standard differential equation solvers of the respective languages. Results indicate that by changing from the tailor-made solvers to using the built-in solvers leads to a speed increase in the order of 6 times. Furthermore, results indicate that the Julia implementations are ca. 5 times faster than the MATLAB implementations. Overall, the fastest Julia implementation was 36 times faster than the current MATLAB implementation. The MATLAB execution can be sped up by using MATLAB Coder to convert the code to efficient C-code which is then used to generate a DLL. DLLs can be executed virtually without overhead from Julia. By measuring the execution time for the C-code/DLL vs. a similar implementation in pure Julia, the pure Julia code is ca. 12% *faster* than the compiled C code.

Next, the possibility of automatic linearization of the population balance model in Julia is studied. This is shown to be relatively straightforward. The linear approximation is very good for an input perturbation of 10%, and relatively good for an input perturbation of 50%. This indicates that it may be possible to use a linear model approximation for control design.

Keywords: linear regression, nonlinear regression, thermal model, machine learning, surrogate model, hybrid model.

1 Introduction

1.1 Background

Industrially produced fertilizers are of key importance in order to produce enough food for a growing global pop-
ulation. Fertilizers in the form of granules is allows for simple application and spreading of fertilizers. The quality of such fertilizers are determined by the average size and the size distribution of the fertilizer, as well as liquid content and porosity. Granulation of fertilizers at times lead to oscillatory operation, which widens the size distribution and increases the energy consumption. It is of interest to develop dynamic population balance models for fertilizer production which describes the size distribution, to some degree explains the product quality, and allows for understanding of what operating conditions lead to oscillatory behavior. Dynamic population balance models are partial differential equations in time as well as external and internal variables. The external variables are spatial position, while the internal variables are particle size, humidity, porosity, etc. The result is that dynamic population balance models are demanding to solve, both numerically and because of the model size.

To be used for on-line production planning/control, it may be necessary to solve the model much faster than real time, e.g., in some types of state estimators and in some types of optimization based control algorithms. It is therefore of interest to explore the possibility of optimizing the model formulation for fast execution within a given computer language, but also to explore whether different languages give different execution time. Some control algorithms may use a linear model approximation. Because of the complexity of the model, it is also of interest to study whether the linearization of the model can be automated in a given language. The ultimate goal of the population balance model is to see whether production with better quality and reduced energy consumption can be achieved.

1.2 Previous work

Population balances describe dynamic systems with both external and internal coordinates, leading to highly distributed models which are time consuming to solve (Wang and Cameron, 2007; Litster and Ennis, 2004; Iveson et al., 2001; Ramkrishna, 2000). (Vesjolaja et al., 2018) describe a population balance model for granulation of fertilizers, including both growth by layering and growth by agglomeration. As a population balance model, the model is relatively simple and homogeneous in the drum axial position as external coordinate, and particle size as internal coordinate. The two growth mechanisms require different types of discretization algorithms. In this simple implementation, the particle size is discretized into 80 different sizes. With 80 states in the model, the key output is the particle

size median. The resulting model is relatively complex as a dynamic model for control, but still simple as pertaining a population balance model. In a future stage, models will be extended with distribution in the external coordinate. It is of interest to compare different modeling languages wrt. solution efficiency.

The model is designed for control synthesis. Standard controllers include proportional (P) and proportional + integral (PI) controllers, which often are tuned based on some tuning rule, (Åstrøm and Murray, 2008). Controllers of mid-level complexity are based on linear approximations of the model, e.g., root locus methods, synthesis based on Nyquist, Nichols, or Bode diagrams, as well as linear quadratic controllers (LQR) and linear Model Predictive Control (MPC), (Maciejowski, 2002). Thus, it is also of interest to consider modeling languages wrt. how they can aid in controller synthesis, e.g., by providing linearized model approximation. Examples of popular languages for solving such models include MATLAB and C++, but a recent, free language such as Julia (Bezanson et al., 2017) with an extensive package for solving differential equations (Rackauckas and Nie, 2017) is an interesting candidate. Julia uses Just-in-Time (JIT) compilation with strong typing, and thus provides a bridge between easy-to-use script languages and compiled languages. Julia also has other advantages with simple-to-use, free packages for Automatic Differentiation (AD) and linearization (Revels et al., 2016), simple-to-use, free packages for computing with distributions (Besanon et al., 2019) such as particle size distribution, etc.

1.3 Overview of the paper

The current model has been implemented in MATLAB with a fixed, user-developed step-length RK4 solver, and a user-developed routine for computing the median of the distribution. We consider to replace the user-developed routines in MATLAB with built-in routines from the ODE solver tools to see if this can make the MATLAB implementation more efficient. Next, we consider implementing the model in Julia and compare the execution speed in Julia vs. that of MATLAB. Both a direct translation of the user-developed RK4 solver and median computation is used, as well as the use of Julia packages such as DifferentialEquations and Statistics. We also compare the execution time of the model written in pure Julia, vs. conversion of the MATLAB model to C-code/DLL using MATLAB Coder from MathWorks. Finally, we study how Julia can be used for linearization of the model.

The paper is organized as follows. In Section 2, an overview of the granulation process is given. In Section 3, key elements of the MATLAB implementation are discussed, with a comparison of the current implementation vs. the use of the built-in ODE solvers. Next, implementation issues for Julia are discussed. Then, simulation results are provided, with a comparison of execution speed. In Section 4, the possibility of automatic linearization of the model using Julia is discussed, with some simple re-

sults. Finally, some conclusions are drawn in Section 5.

2 Overview of Industrial Granulation

2.1 Fertilizer granulation

Granulation processes are used in a wide range of industrial applications, such as pharmaceutical and fertilizer industries. The research reported here is focused on the last part of NPK (Nitrogen, Phosphorus, Potassium) fertilizer production. A granulation loop is used to produce different grades, i.e., various N:P:K ratios, of fertilizers. The NPK fertilizer is a high value type of fertilizer containing the three main elements essential for crop nutrition. Various NPK grades are specially developed for different crops growing in different climates and soils.

2.2 Granulation loop

A typical schematic of a granulation process with a recycle loop is shown in Figure 1.

The granulation loop consists of a granulator, a granule classifier (screens), and a double-roll crusher. During the granulation process, a slurry of liquid ammonium nitrate and partly dissolved minerals is solidified to form granules. Granules that are too small (under-sized particles) are recycled to the granulation unit, while granules that are too large (over-sized particles) are crushed and then recycled back to the granulator. Different granulation mechanisms are responsible for the granule formation in the granulator, depending on the granulator type and operational conditions. The granulator can be of different types, e.g., a spherodizer, a rotary drum granulator, a fluidized bed granulator, a pan granulator, etc. Some of the granulation mechanisms that are responsible for the granule formation are particle growth due to layering and particle agglomeration. Particle growth due to layering is a continuous process during which particle growth occurs due to a successive coating of a liquid phase onto a granule. Binary particle agglomeration is a particle growth mechanism that occurs due to successful collision of two particles, resulting in the formation of a larger, composite particle (Litster and Ennis, 2004; Vesjolaja et al., 2018).

2.3 Production challenges

Fertilizer manufacturing using the granulation process has received considerable research interest during the last few decades, due to (i) the increasing requirements for efficient production of high quality fertilizers for increased food production in a growing global population, (ii) difficult process control and operation. Process control of granulation loops is challenging since the particle size distribution (PSD) of the granules leaving the granulator is wider than the required PSD of the final product. A typical recycle ratio between the off-spec particles and the required product-sized particles is 4:1 respectively. In addition, granulation loops may show oscillatory behavior for certain operating points. Typical oscillations seen in an-industrial scale fertilizer granulation plants is depicted
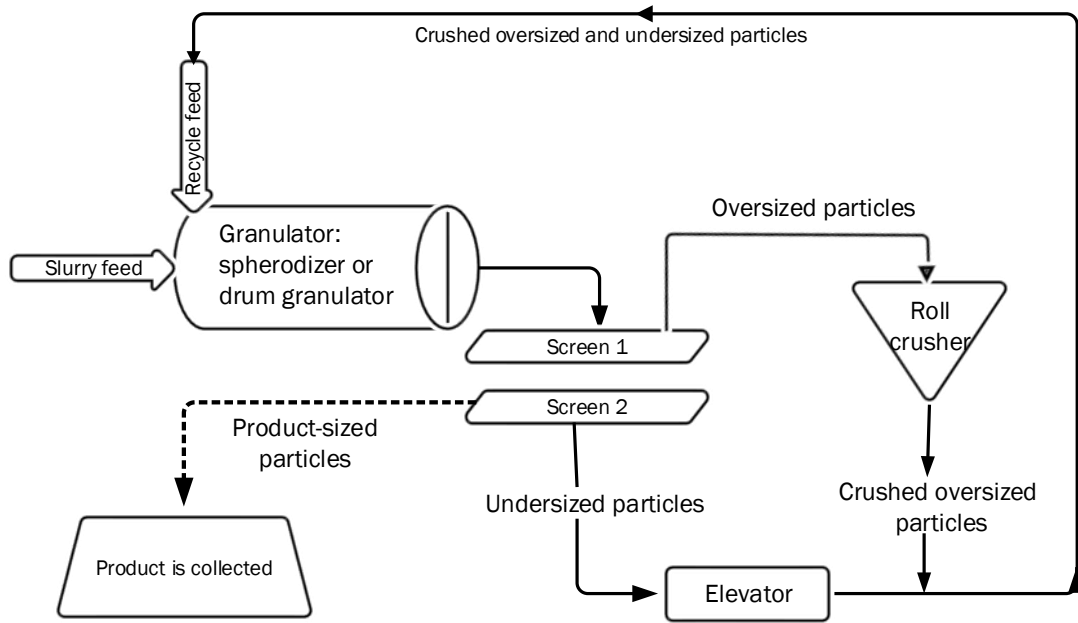
Figure 1. Granulation loop used in fertilizer industry.
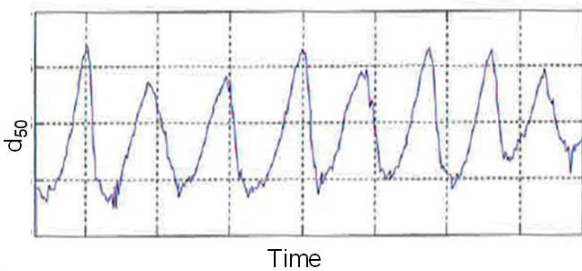
Figure 2. Oscillatory behavior observed in PSD (measured as d50) of the produced granules in fertilizer industry.

in Figure 2.

Control and operation of a granulation loop is still challenging. Some research papers with focus on process control of the granulation process are (Buck et al., 2016; Ramachandran and Chaudhury, 2012; Herce et al., 2017; Ramachandran et al., 2009; Valiulis and Simutis, 2009; Wang et al., 2006; Cameron et al., 2005).

2.4 Problem limitation

Here, we consider a model of the drum granulator with both layering and agglomeration, but we do not include the screening or the recycling. The main purpose here is to study the efficiency of model formulation and solution in MATLAB vs. Julia, and the possibility to use modern modeling tools for control analysis.

3 Model implementation details

3.1 Overview of model

The population balance for combined layering and agglomeration is discussed in (Vesjolaja et al., 2018), and is for form

$$\frac{\partial n(L,t)}{\partial t} = -\frac{\partial}{\partial L}(G \cdot n(L,t)) + B(L,t) - D(L,t) + \dot{n}_i \gamma_i - \dot{n}_e \gamma_e$$

(1)

where $n(L,t)$ is the number density as a function of particle diameter L and time t, G is the growth rate relevant for layering, $B(L,t)$ is the birth rate relevant for agglomeration, while the death rate $D(L,t)$ describes particle disintegration. n_i is the influent number flow rate, γ_i is the influent size distribution, while n_e and γ_e are similar quantities for the effluent. Here, perfect (external) mixing has been assumed in the granulator drum. Alternatively to a number population balance as in Eq. 1, it is often more convenient to describe the mass population balance, where m is related to n via

$$n = \frac{6m}{\pi \rho L^3}.$$

(2)

When discretizing the particle size space in N_p particle sizes, the mass population balance has the following form:

$$\frac{dM_{1:N_p}}{dt} = f_{agg}(M_{1:N_p};\theta) + f_{growth}(M_{1:N_p},\dot{M}_{sl};\theta) + \dot{M}_i \gamma_i - \dot{M}_e \gamma_e,$$

where $M_{1:N_p}$ is the vector of masses within the N_p size ranges $L_1,\ldots,L_{N_p}$, θ is some model parameter, $\dot{M}_{sl}$ is the

feed rate of slurry fertilizer spray, and $\dot{M}_e$ is given by some expression. The quality of interest for control is the *median* of the particle size distribution (PSD) in the effluent, measured in particle diameter and denoted by $d_{50,e}$,

$$d_{50,e} = \text{median}\left(L, M_{1:N_p}\right).$$

Introducing $x = M_{1:N_p}, u = \left(\dot{M}_i, \dot{M}_{sl}\right)$, $y = d_{50,e}$, we can express the model in the more abstract form

$$\frac{dx}{dt} = f\left(x, u, t; \theta\right)$$
$$y = g\left(x, u, t; \theta\right).$$

For the work reported in (Vesjolaja et al., 2018), the mass-based model was implemented in MATLAB using a tailor-made, fourth order Runge-Kutta (RK4) fixed step-length solver with integral computation of $d_{50,e}$ including conversion from mass to diameter, and storage of the output $d_{50,e}$.

3.2 MATLAB

It is of interest to rewrite the MATLAB code of (Vesjolaja et al., 2018) and use the built-in MATLAB ODE solvers, to see if the code can be made faster. Some advantages of MATLAB are a rich toolbox suite, easy debugging, and excellent documentation.

ODE solvers in MATLAB have the following call structure:

```
1  [t , x] = solver (odefun, tspan, x0, options)
```

where we considered the following `solver` alternatives: standard solvers **ode45**, **ode23**, and ode113, as well as stiff solvers ode15s, ode23t, and ode23tb. When solved as ordinary differential equations, these solvers only store the state, $M_{1:N_p}$, thus the output $d_{50,e}$ must be computed by post processing. It is possible to solve differential algebraic equations (DAEs) with MATLAB, e.g., introducing a singular mass matrix, but such a reformulation is somewhat clumsy for certain types of problems (e.g., with re-circulation), and DAE solver are slower than ODE solvers.

MATLAB has decent support for computing quantiles of distributions, e.g., $d_{50,e}$ of the particle size distribution (PSD). However, because of the conversion from mass distribution to diameter distribution indicated in Eq. 2, support for quantile computation with weighting is required; this is not supported in standard MATLAB. Because of this, the original function for computing $d_{50,e}$ from the distribution $M_{1:N_p}$ is preserved.

3.3 Julia

In order to compare how fast the model can be solved, it is of interest to compare the MATLAB implementation with an implementation in another language. Standard reference languages for speed are C/C++ and FORTRAN. However, the relatively new language Julia which is a dynamic language in style, while using Just-in-time compilation for execution, is also known to be fast — at least

when properly implemented. Julia has excellent packages for differential equation solvers, and very good support for statistics.

In general, Julia has relatively rudimentary documentation for packages compared to MATLAB toolboxes. Some integrated development environments (IDE) with debugger are starting to appear, but they are still inferior to the MATLAB IDE. However, packages are, in general, of good–excellent quality.

For comparison with MATLAB, first the tailor-made RK4 solver used in (Vesjolaja et al., 2018) was translated more or less directly to Julia. Next, just like for MATLAB, a rewrite of the code was made to take advantage of the differential equation solvers in Julia's package `DifferentialEquations.jl`. Every ODE solver in MATLAB has an equivalent solver algorithm in Julia (https://docs.sciml.ai/stable/solvers/ode_solve). Specifically, Julia versions of the non-stiff MATLAB solvers were used: DP5 (**ode45**), BS3 (**ode23**), VCABM (ode113). Because experiments with MATLAB indicated that the system is non-stiff, stiff Julia solvers were not considered. Finally, the standard Julia solver Tsit5 was used.

Julia has support for quantile computations with weights, and in the rewritten code, this function was used to compute $d_{50,e}$.

3.4 C-code/DLL

MATLAB has tools for automatic conversion of code to C-code and compilation into DLLs. Such DLLs can then be included in Simulink blocks. It is also possible to call DLLs from Julia virtually without overhead. This makes it possible to compare the execution time of C-code vs. Julia and MATLAB.

3.5 Comparison

In this section, two implementations of the granulation drum model in each programming language will be compared, i.e., 4 implementations (with some variation in solvers). The tailor-made RK4 solvers (MATLAB and Julia) will be referred to as TM-RK4. Next, we rewrite the tailor made code to be in the standard forms for use with built-in ODE solvers. For these, we refer to the code with language name and solver, e.g., M-**ode45** for solver **ode45** in MATLAB, and J-DP5 for the similar DP5 solver for Julia.

Table 1 shows a comparison of run-time (execution time) for MATLAB and Julia for simulation over 1.5 h. The results displayed are the best run out of 20 runs, with $N_p = 80$. For the tailor-made solvers, a step size of $h = 10\,\text{s}$ is used. Both for MATLAB and Julia, adaptive time stepping is used in the built-in ODE solvers. In general, Julia appears to use longer step-length and thus have fewer steps. This could be because most Julia solvers include interpolation of the solution for improved accuracy.

Finally, the model implementation with tailor-made RK4 solver (TM-RK4) was converted to C-code/a DLL using MATLAB Coder from MathWorks. This DLL re-

Table 1. Comparison of run time for the various implementations. TM: tailor-made, M: MATLAB, J: Julia. Columns for MATLAB and Julia show absolute simulation time in seconds, and in parenthesis: computation time relative to the fastest combination. Finally, the right-most column shows the relative run-time of MATLAB vs. Julia. In all solvers except the fixed step-length TM-RK4 algorithm, absolute tolerance 10^{-6} and relative tolerance 10^{-3} was used.

Algorithm	MATLAB	Julia	MATLAB/Julia
TM-RK4	171.2 s (36)	40.3 s (8.6)	4.25
M-ode45, J-DP5	29.4 s (6.3)	6.2 s (1.3)	4.74
M-ode23, J-BS3	40.7 s (8.7)	6.0 s (1.3)	6.78
M-ode113, J-VCABM	27.4 s (5.8)	4.7 s (1.0)	5.83
M-ode15s	59.3 s (13)	–	–
M-ode23t	63.9 s (14)	–	–
M-ode23tb	88.6 s (19)	–	–
J-Tsit5	–	6.9 s (1.5)	–

turned $f(x,t)$ in the ODE $\frac{dx}{dt} = f(x,t)$ and was then called virtually without overhead from Julia in an Euler integration loop, and the execution speed was compared to the TM-RK4 method implemented in Julia and called in an identical Euler integration loop. In this case, the pure Julia implementation was ca. 12% *faster* than the C-code/DLL. To have a 100% fair speed comparison, the model should have been implemented 100% in C and Julia by experts in the respective languages. However, the for loop in Julia for doing Euler integration is efficient, with most of the computational load taking place in computing $f(x,t)$. Because of this, we believe that execution speed in Julia is relatively similar to what can be achieved in C for this type of problem.

4   Model linearization in Julia

For model-based control design, a linear approximation of a model is often sought in form

$$\frac{dx}{dt} = Ax + Bu$$
$$y = Cx + Du$$

where matrices A, B, C, and D are found as the following Jacobians at the operating point $*$ given by (u^*, x^*):

$$A = \left. \frac{\partial f(x,u,t;\theta)}{\partial x} \right|_*$$
$$B = \left. \frac{\partial f(x,u,t;\theta)}{\partial u} \right|_*$$
$$C = \left. \frac{\partial g(x,u,t;\theta)}{\partial x} \right|_*$$
$$D = \left. \frac{\partial g(x,u,t;\theta)}{\partial u} \right|_*.$$

The linear approximation is believed to give good approximation to the nonlinear model as long as the perturbations in the system are "small" relative to the operating point given by (u^*, x^*).

For complex models, it has traditionally been laborious to develop a linear approximation. However, some modern languages has support for Automatic Differentiation for exact linearization. Commercial language MATLAB has support for this, but since Julia is a free language, it is of interest to see how this can be done in Julia. Julia has several packages for carrying out automatic differentiation; here we use package `ForwardDiff.jl`. Assume that we have found a steady operating point (u^*, x^*). Next, we formulate specialized models for the vector fields $f(x,u)$ and $g(x,u)$:

$$f_x(x) : x \to f(x, u^*)$$
$$f_u(u) : u \to f(x^*, u)$$
$$g_x(x) : x \to g(x, u^*)$$
$$g_u(u) : u \to g(x^*, u).$$

By associating `f_x`$= f_x$, `f_u`$= f_u$, `g_x`$= g_x$, `g_u`$= g_u$, as well as `x_ast`$= x^*$ and `u_ast`$= u^*$, we can compute matrices A, B, C, D as follows:

```
1  using ForwardDiff
2  A = ForwardDiff.jacobian(f_x, x_ast);
3  B = ForwardDiff.jacobian(f_u, u_ast);
4  C = ForwardDiff.jacobian(g_x, x_ast);
5  D = ForwardDiff.jacobian(g_u, u_ast);
```

We can now compare transients of the nonlinear model and the linear model from a chosen steady state, and compare, e.g., the final time mass distribution $M_{1:N_p}$ in the effluent, and the output $d_{50,e}$. Figure 3 shows the comparison with a 10% increase in the input u.

Figure 4 shows the comparison with a 50% increase in the input u.

Figure 3 shows only 0.08 % error in $d_{50,e}$ in the linear approximation for 10% change in the input u, which is acceptable for many applications. Figure 4 shows 1.58 % error in $d_{50,e}$ in the linear approximation for 50% change in the input u, which perhaps is a large error for prediction purposes, but even with this error, the indicated response is not too different from the nonlinear model, and this error can probably be compensated for by feedback control.

5   Conclusions

The outset of this study was an interest in comparing the run-time for differential equation solvers for different lan-
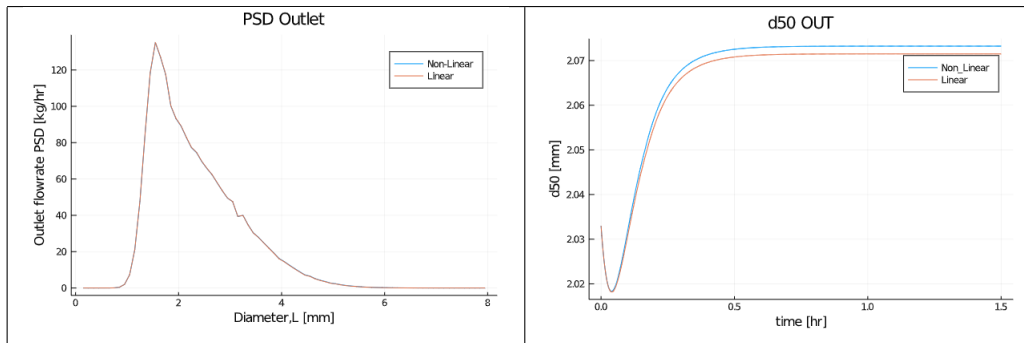
Figure 3. Comparison between linear and nonlinear models where both inputs increase 10%.
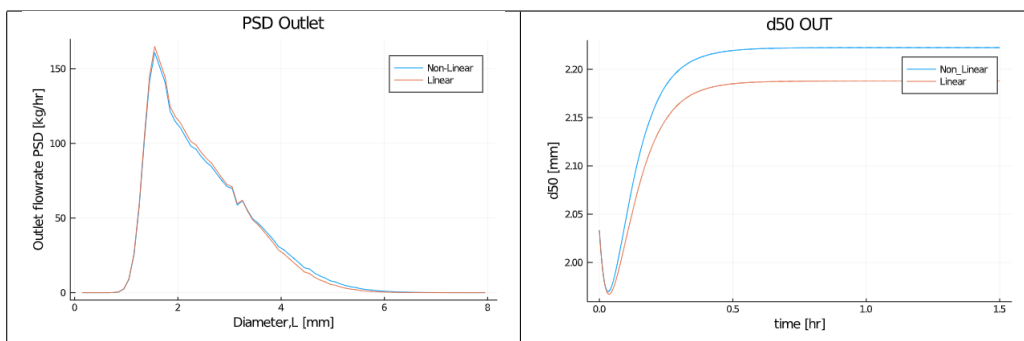
Figure 4. Comparison between linear and nonlinear models where both inputs increase 50%.

guages and different implementations. The test model is a population balance model for the granulation drum of fertilizer production, i.e., without product recycling. Such models are of high order, and are relatively complex to solve numerically. In order to use such a model in optimization based control algorithms, it must be possible to solve the model much faster than real time, so execution time is an issue.

The original implementation of the model under study was in MATLAB, and used a tailor-made implementation of a 4th order Runge-Kutta solver in order to store/present on-line the output, which is the median of the particle diameter distribution. This implementation was transferred to Julia for comparison.

In an extension of this rewrite to Julia, the MATLAB code was rewritten from the tailor-made solver implementation to a form which can use the standard MATLAB ODE solvers. Likewise, the Julia code was rewritten to take advantage of the Julia solvers in package DifferentialEquations.jl. In both cases, every attempt was made to make the code efficient, e.g., taking advantage of vectorization in MATLAB.

Speed comparisons indicate that Julia is typically ca. 5 times faster than MATLAB. Similarly, the results indicate that utilizing the built-in solvers in the languages is in the order of 6 times faster than using the tailor-made solvers. To this end, the fastest Julia implementation is approximately 36 times faster than the original, tailor-made MATLAB implementation.

Some experiments have been made with Julia DAE solvers instead of the ODE solvers. The advantage with DAE solvers is that one can compute outputs on-line instead of by post processing. Although a thorough comparison has not been carried out, initial attempts indicate that using DAE solvers approximately doubles the computation time compared to ODE solvers.

Some initial attempts have also been made with implementing the tailor-made solver in C. Specifically, the MATLAB code was converted to C-code/DLL using MATLAB Coder. The resulting DLL can be called from Julia virtually without overhead, and the execution speed was compared to that of the pure Julia code. The result of this comparison was that Julia was ca. 12% *faster* than the C-code implementation of the model.

Overall, the comparison between the three languages MATLAB, Julia, and C indicate that for solving ODEs, Julia and C have relatively similar execution speed, in spite of Julia being a scripted language — although with Just in Time Compilation. Execution in both Julia and C is considerably faster than in MATLAB. Using MATLAB Coder makes it possible to regain the speed advantage of Julia; however, Julia and the Julia eco-system is free.

In addition to speed comparisons, we have checked the possibility of automatic linearization of the Julia code. This can be done using available, free Julia packages, and appeared to be relatively straightforward, with the linear approximation of the granulation model being quite close to the nonlinear solution in realistic cases. This indicates that the linear approximation probably suffices for control design. It is also possible to do such automatic lineariza-

tion in MATLAB. However, because Julia is a free tool while MATLAB is an expensive, commercial tool, exploring that possible is perhaps less interesting.

Future work will focus on implementing Julia and MATLAB code for the extended system with product screening and re-circulation, as well as automatic computation of a linear approximation for control design. In a longer horizon, it is also of interest to consider possibilities for solving the model over a distributed computer network, or on GPUs.

References

Karl Johan Åstrøm and Richard M. Murray. *Feedback Systems. An Introduction for Scientists and Engineers*. Princeton University Press, Princeton, NJ, 2008. ISBN 978-0-691-13576-2.

Mathieu Besanon, David Anthoff, Alex Arslan, Simon Byrne, Dahua Lin, Theodore Papamarkou, and John Pearson. Distributions.jl: Definition and modeling of probability distributions in the juliastats ecosystem. *arXiv*, 2019.

Jeff Bezanson, Alan Edelman, Stefan Karpinski, and Viral B. Sha. Julia: A Fresh Approach to Numerical Computing. *SIAM Review*, 49(1):65–98, 2017. doi:10.1137/141000671.

A. Buck, R. Durr, M. Schmidt, and E. Tsotsas. Model predictive control of continuous layering granulation in fluidized beds with internal product classification. *Journal of Process Control*, 45:66–75, 2016.

I.T. Cameron, F.Y. Wang, C.D. Immanuel, and F. Stepanek. Process systems modelling and applications in granulation: A review. *Chemical Engineering Science*, 60:3723–3750, April 2005.

C. Herce, A. Gil, M. Gil, and C. Cortés. A cape-taguchi combined method to optimize a npk fertilizer plant including population balance modeling of granulation-drying rotary drum reactor. *Computer Aided Chemical Engineering*, 40:49–54, 2017.

S.M. Iveson, J.D. Litster, K. Hapgood, and B.J. Ennis. Nucleation, growth and breakage phenomena in agitated wet granulation processes: a review. *Powder Technology*, 117(1–2): 3–39, 2001.

Jim Litster and Bryan Ennis. *The Science and Engineering of Granulation Processes*, volume 15 of *Particle Technology Series*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2004. ISBN 1-4020-1877-0.

Jan M. Maciejowski. *Predictive Control with Constraints*. Prentice Hall, Harlow, England, 2002.

Christopher Rackauckas and Qing Nie. DifferentialEquations.jl — A Performant and Feature-Rich Ecosystem for Solving Differential Equations in Julia. *Journal of Open Research Software*, 5(15), 2017. doi:10.5334/jors.151.

R. Ramachandran and A. Chaudhury. Model-based design and control of a continuous drum granulationprocess. *Chemical Engineering Research and Design*, 90(8):1063–1073, 2012.

R. Ramachandran, C.D. Immanuel, F. Stepanek, J.D. Litster, and F.J. Doyle, III. A mechanistic model for breakage in population balances of granulation: Theoretical kernel development and experimental validation. *Chemical Engineering Research and Design*, 87(4):598–614, 2009.

D. Ramkrishna. *Population Balances. Theory and Applications to Particulate Systems in Engineering*. Academic Press, London, 2000.

J. Revels, M. Lubin, and T. Papamarkou. Forward-Mode Automatic Differentiation in Julia. *arXiv:1607.07892 [cs.MS]*, 2016. URL https://arxiv.org/abs/1607.07892.

G. Valiulis and R. Simutis. Particle growth modelling and simulation in drum granulator-dryer. *Information Technology and Control*, 28(2), 2009.

Ludmila Vesjolaja, Björn Glemmestad, and Bernt Lie. Population balance modelling for fertilizer granulation process. In Lars Erik Øi, Tiina Komulainen, Robin T. Bye, and Lars O. Nord, editors, *Proceedings of the 59th Conference on Simulation and Modelling*, pages 95–102, Oslo Metropolitan University, Oslo, Norway, September 2018. SIMS, Linköping University Electronic Press. doi:http://doi.org/10.3384/ecp181531.

F.Y. Wang and Ian T. Cameron. A multi-form modelling approach to the dynamics and control of drum granulation processes. *Powder Technology*, 179(1–2):2–11, 2007.

F.Y. Wang, X.Y. Ge, N. Balliu, and I.T. Cameron. Optimal control and operation of drum granulation processes. *Chemical Engineering Science*, 61(1):257–267, 2006.

Heat and Mass Transfer Model for Droplets with Internal Circulation

Mathias Poulsen   Kim Sørensen   Thomas Condra

Department of Energy Technology, Aalborg University, Denmark, {mps,kso,tc}@et.aau.dk

Abstract

In large droplets the internal resistance to heat or mass transfer has to be accounted for. Two respective models for low and high Reynolds numbers are investigated. The models are solved numerically for a range of modified Peclet numbers and the corresponding transfer numbers, representing either the Nusselt or Sherwood numbers, are determined. The results for each model are fitted to produce an expression that can be easily evaluated for use in a CFD code. The fits has mean deviations of 0.63% and 0.035% for the low and high Reynolds number models respectively. A proposed switching Reynolds number is used to combine the models and the combined model is compared to temperature measurements of free falling water droplets. It was found that the model is in good agreement with the data for the smallest droplets whereas it deviates as much as 40% for the larger droplets in the data set.

Keywords: internal circulation, mass transfer, heat transfer, droplets, CFD

1 Introduction

The droplet diameter in sprays with a large mass flow are typically in the order of 10^{-3} - 10^{-5} m. When modelling heat and mass transfer for small droplets, the transfer coefficients inside the droplets can be assumed to be infinite, and the overall transfer coefficients can be determined externally, by utilising correlations such as the Frössling-Marshall equation (Bird et al., 2002). However, in larger droplets the heat and mass transfer coefficient is significantly lower and this must be taken into consideration.

Processes involving sprays are often complex two-phase flow with many different types of interactions between the gas and liquid. Computational fluid dynamics (CFD) is an exceptional tool for capturing the respective gas-liquid interactions. As CFD requires relatively large computational capacity, the models used to describe the gas-liquid interaction should be fairly simple to evaluate, so as to limit the computational resources necessary.

2 Governing Equations

When modelling a two phase flow, the transfer of heat and mass are expressed as functions of intrinsic phase properties, that being temperature or mass concentration. The flux across the droplet interface can be described from both the gas and droplet side as:

$$j = \kappa_g \left(\phi_g - \phi_g^{if} \right) \text{ (gas side)} \tag{1}$$

$$j = \kappa_d \left(\phi_d^{if} - \phi_d \right) \text{ (droplet side)} \tag{2}$$

where j is the flux across the interface, ϕ is the driving potential, e.g. mass concentration or temperature of the phase on the gas or droplet side, ϕ^{if} is the potential at the interface and κ is the transfer coefficient on each side of the interface.

For larger droplets the internal movement of liquid causes a non-uniform distribution of the transfer quantity, which in turn causes the transfer coefficient to vary depending on the distribution of mass or heat. To describe the heat or mass transport inside the droplet a continuity equation can be expressed for the quantity of interest as:

$$\frac{\partial \phi}{\partial t} + \nabla u \phi = \nabla \left(\Gamma \nabla \phi \right) \tag{3}$$

where u is the internal droplet velocity, and Γ is the mass or thermal diffusion coefficient.

In the analysis below, the following assumptions are made:

A. 1 The droplet is spherical

A. 2 The fluid of the droplet is incompressible and viscous

A. 3 Fluid properties are constant throughout the droplet and independent of ϕ

A. 4 The distribution of ϕ and velocity field is axisymmetric

A. 5 The velocity field time independent

A. 6 The droplet interface is at equilibrium with the gas phase

2.1 Low Reynolds Model

At low Reynolds numbers the streamlines described by (Hadamard, 1911) can be used to calculate the velocity field within the droplet. The streamlines are shown in the right half of Figure 1. The streamline function in spherical coordinates can be given as:

$$\psi(r,\theta) = \frac{1}{4} u_{rel} a^2 \sin^2(\theta) \left(\frac{\mu_g}{\mu_g + \mu_d} \right) \left(\frac{r}{a} \right)^2 \left[1 - \left(\frac{r}{a} \right)^2 \right] \tag{4}$$

where r and θ are the coordinates with origin at the centre of the droplet, u_{rel} is the relative velocity between the droplet and gas, μ is the dynamic viscosity of the gas and droplet respectively and a is the radius of the droplet.

The respective velocity components can be derived from the stream function in Equation 4 by (Munson et al., 2013):

$$u_r = \frac{1}{r^2 \sin(\theta)} \frac{d\psi(r,\theta)}{d\theta} \quad (5)$$

$$u_\theta = -\frac{1}{r \sin(\theta)} \frac{d\psi(r,\theta)}{dr} \quad (6)$$

This approach has previously been investigated by Wellek et al. (1970), where the continuity equation presented in Equation 3 is rearranged to make it dimensionless.
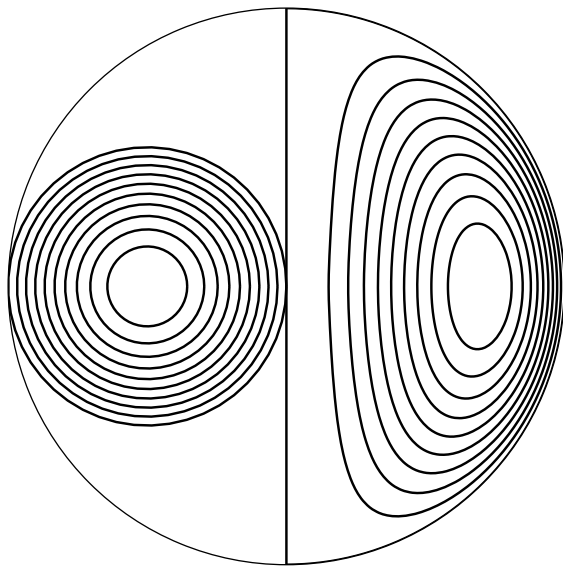
Figure 1. Streamlines inside the droplet. On the left side is the tori streamlines for the high Reynolds model and on the right side is the Hadamard streamlines for the low Reynolds model.

The dimensionless transport equation in spherical coordinates is given as:

$$\frac{\partial \Phi}{\partial \tau} + \frac{1}{4} \left[Pe' \left(1 - R^2\right) \cos(\theta) - \frac{2}{R} \right] \frac{\partial \Phi}{\partial R} - \frac{\partial^2 \Phi}{\partial R^2}$$
$$+ \frac{1}{4} \left[Pe' \frac{\left(2R^2 - 1\right)}{R} - \frac{\cot(\theta)}{R^2} \right] \frac{\partial \Phi}{\partial \theta} - \frac{1}{R^2} \frac{\partial^2 \Phi}{\partial \theta^2} = 0 \quad (7)$$

$$\tau = \frac{t\Gamma}{(2a)^2} \quad (8)$$

$$Pe' = \left(\frac{\mu_g}{\mu_g + \mu_d} \right) Pe \quad (9)$$

$$Pe = \frac{u_{rel}(2a)}{\Gamma} \quad (10)$$

$$R = \frac{r}{a} \quad (11)$$

$$\Phi = \frac{\phi(t) - \phi(0)}{\phi(\infty) - \phi(0)} \quad (12)$$

Where τ is the Fourier number, Pe' is the modified Peclet number, R is the relative radial position and Φ is the normalised mass or heat.

By making Equation 3 dimensionless a solution for a given Pe' can be used independent of the specie or liquid in question.

For the limiting cases where the transfer is dominated by either diffusion or convection i.e. $Pe = 0$ or $Pe = \infty$, there are analytical solutions. These are expressed as volume averages of the quantity of interest over the droplet, which can be given as:

$$\bar{\Phi} = \int_V \Phi \, dV \quad (13)$$

For the case of pure diffusion the solution of Equation 7 can be given as (Newman, 1931):

$$\bar{\Phi} = 1 - \frac{6}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} \exp\left(-n^2 \pi^2 \tau\right) \quad (14)$$

where $\bar{\Phi}$ is the volume averaged quantity of interest and V is the droplet volume.

For the other case where the transfer is dominated by convection a solution was found by separation of variables (Kronig and Brink, 1951). This method results is an infinite sum where each addition requires a new coefficient and an eigenvalue. In the original work only one set of coefficients and eigenvalues was found. Later the number of coefficients and eigenvalues where extended (Oliver and Souccar, 2006). The solution can be given as:

$$\bar{\Phi} = 1 - \frac{3}{8} \sum_{n=1}^{\infty} A_n^2 \exp\left(-16\lambda_n \tau\right) \quad (15)$$

where A_n and λ_n is the respective coefficient and eigenvalue to the nth addition to the sum.

2.2 High Reynolds model

For a droplet at high Reynolds number a model based on the streamlines being replaced by a system of tori is used (Handlos and Baron, 1957), see the left half of Figure 1. In this model the transfer inside the droplet is based on eddy diffusion, which is expressed as function of the relative radial position inside the tori system. As only the tori are responsible for the transfer mechanism, the volume that contributes to transfer is 59% of the entire droplet.

The transfer equation for the high Reynolds model are given as:

$$\frac{\partial \Phi}{\partial \tau} = \frac{Pe'}{128} \frac{1}{\zeta} \frac{\partial}{\partial \zeta} \left(\left(6\zeta^3 - 8\zeta^2 + 3\zeta\right) \frac{\partial \Phi}{\partial \zeta} \right) \qquad (16)$$

$$(17)$$

where ζ is the coordinate centred at the tori.

A solution for the high Reynolds model was given by (Olander, 1966), but this is only valid for $Pe'\tau/128 > 0.1$. The solution presented by (Olander, 1966) is given as:

$$\bar{\Phi} = 1 - 0.64 \exp \left(-\frac{2.80 Pe'\tau}{128} \right) \qquad (18)$$

With the limitation of Equation 18, it will not be used directly further analysis, but it will still be used to validate the implementation of the model.

2.3 Boundary Conditions

For the boundary at the surface of the droplet, assumption A. 6 gives the following condition:

$$\Phi(1, \tau) = 1 \qquad (19)$$

For the low Reynolds model the boundary condition at the droplet halves can be given as:

$$\frac{\partial \Phi(R, 0, \tau)}{\partial \theta} = \frac{\partial \Phi(R, \pi, \tau)}{\partial \theta} = 0 \qquad (20)$$

2.4 Transfer coefficient

The classical transfer models for heat or mass are typically correlated to each of their respective dimensionless group which can be expressed as:

$$N_t = \frac{\kappa_d(2a)}{\Gamma} \qquad (21)$$

where N_t is the Sherwood or Nusselt number.

To determine the dimensionless group for the quantity of interest, it has to be correlated to $\bar{\Phi}$. By defining the dimensionless flux across the droplet interface and rewriting Equations 2 and 21 to express the transfer coefficient and an expression for N_t can be found (Andoe, 1968):

$$J = \frac{2}{3} \frac{d\bar{\Phi}}{d\tau} \qquad (22)$$

$$\kappa_d = \frac{j}{\phi_{p,d}^{if} - \phi_{p,d}} = \frac{J}{1 - \bar{\Phi}} \frac{2a}{\Gamma} \qquad (23)$$

$$N_t = \frac{J}{1 - \bar{\Phi}} \qquad (24)$$

Following the method presented in (Kronig and Brink, 1951) and (Handlos and Baron, 1957), N_t is evaluated for a given Pe', when $\bar{\Phi} = 1 - e^{-1} \approx 0.632$

3 Computational Methodology

Due to the difference in the nature of the governing equations, each model is solved using a different method. For the low Reynolds model the solver scalarTransportFoam from the open source package OpenFOAM is used to solve Equation 7 (Weller et al., 1998). OpenFOAM is written for Cartesian coordinates, therefore Equation 7 is rewritten into the form of Equation 3. Here the velocity components are given as:

$$u_x = u_r \cos(\theta) - R u_\theta \sin(\theta) \qquad (25)$$

$$u_y = u_r \sin(\theta) + R u_\theta \cos(\theta) \qquad (26)$$

The discretisation scheme used for the divergence terms is second order upwind, and the time scheme is Crank-Nicholson.

A structured axis symmetric mesh is created for the droplet half where the centre axis of the droplet is the symmetry axis. Three different meshes with increasing mesh densities of 2400, 5400 and 9600 cells are investigated for varying Pe' to ensure that the solutions is independent of the mesh density. The time step for the different meshes where handled by defining a maximum allowable Courant number of 0.5. In Figure 2 the results for three different mesh densities are shown. It can be seen that there is no significant difference between solution on the different meshes. With a relative difference below 0.1% between
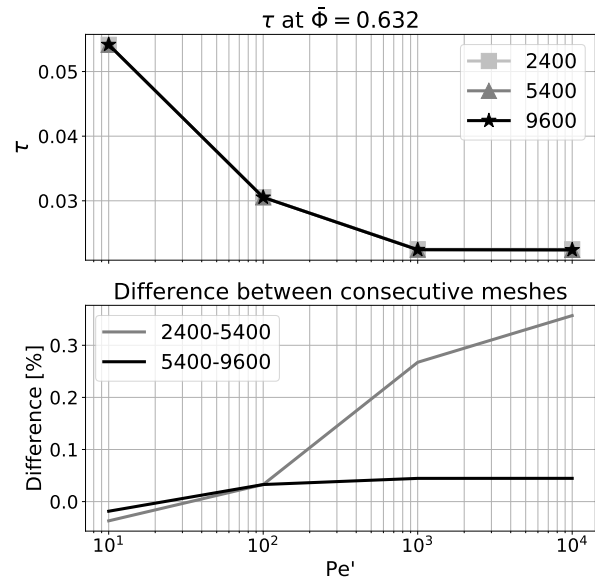
Figure 2. The top plot shows the dimensionless time for the three different mesh densities at different modified Peclet numbers. The bottom plot shows the relative difference between the consecutive meshes.

the mesh with 5400 and 9600 cells, the mesh with 5400 cells are used for further analysis.

For the high Reynold model, Equation 16 is discretised using a second order accurate central finite difference approach. The first and second derivatives are expressed as:

$$\frac{\partial \Phi}{\partial \zeta} \approx \frac{\Phi_{+1} - \Phi_{-1}}{\Delta \zeta} \tag{27}$$

$$\frac{\partial^2 \Phi}{\partial \zeta^2} \approx \frac{\Phi_{+1} - 2\Phi + \Phi_{-1}}{\Delta \zeta^2} \tag{28}$$

For the dimensionless time a forward Euler approximation is used.

Just as for the low Reynolds model, an independence test was done to ensure that the solution is independent of the discretisation. In Figure 3 the value of τ when $\bar{\Phi} = 0.632$ is shown for different values of the modified Peclet number Pe'.
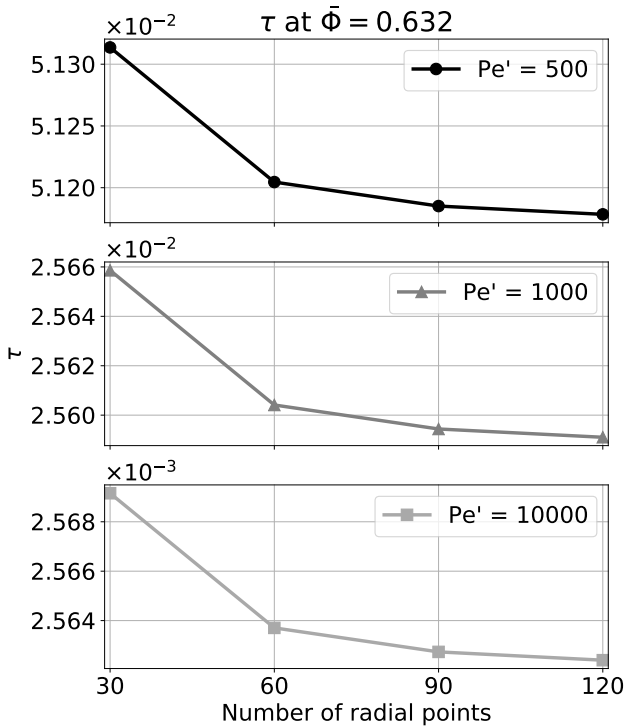
Figure 3. Independence test for the high Reynold model. The value of the dimensionless time τ when $\bar{\Phi} = 0.632$ is shown at varying modified Peclet numbers for different number of radial points used to discretise the droplet.

With a relative difference of 0.013 % or less, between the solution with 120 and 90 radial point for the investigated Pe', the discretisation with 90 radial points are used for further analysis. A study of the dimensionless time step was also done, and it was found that a value of 1×10^{-6} was adequate for all the modified Peclet numbers.

3.1 Model Comparison

As a final validation of the computational methodology, the numerical results from both models are compared

with Equations 14, 15 and 18. The solution for the low Reynolds model at $Pe' = \infty$ is approximated by setting $Pe' = 10^4$. The comparison for the high Reynolds model are shown for $Pe' = 500$.
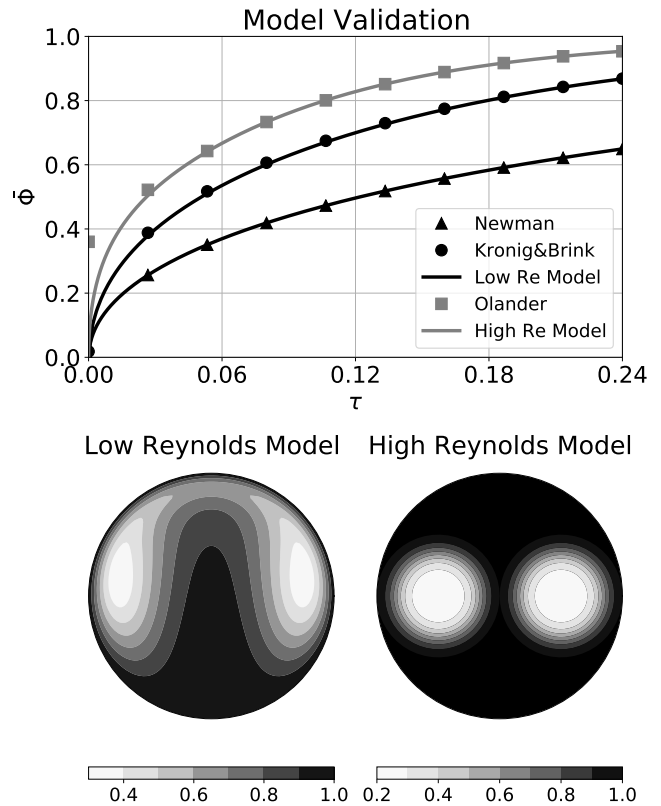
Figure 4. Comparison of the low and high Reynolds model with known solutions to their respective set of equations. On the bottom of the figure, the contours for the low and high Reynolds model are shown for $\bar{\Phi} = 0.632$.

The output from the low Reynolds model at $Pe' = 0$ and $Pe = \infty$ agrees well with solution of Equations 14 and 15, respectively. As stated by (Olander, 1966) the expression in Equation 18 is only accurate for $Pe'\tau/128 > 0.1$, which is why the first point of comparison deviates from the model results, whilst for the remaining points it agrees with model.

4 Results

The models for low and high Reynold number are run over a span of modified Peclet numbers ranging from 0 to 10^4 and the transfer number is evaluated for each modified Peclet number. The resulting transfer numbers are shown in Figure 5:

To get an expression for the transfer numbers as a function of Pe' the two curves shown in Figure 5 are fitted using least-squares. The curve for the low Reynolds model transfer numbers resembles that of a step response to an under-damped second-order system shifted by an amount equal to that of the transfer number of the solution presented by (Newman, 1931), when $Pe' \leq 1000$.
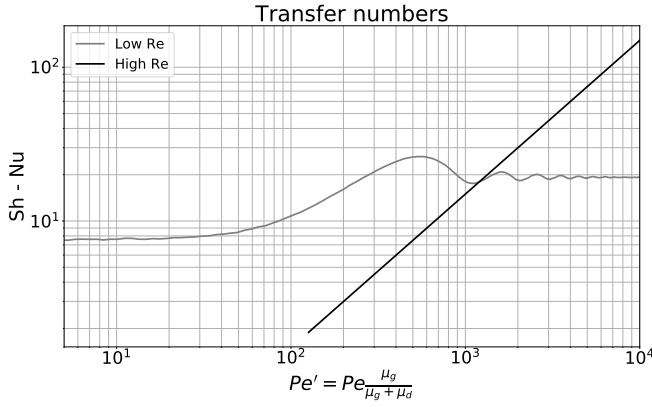
Transfer numbers

Figure 5. Resulting transfer numbers for the low and high Reynold model at varying modified Peclet numbers.

For $Pe' > 1000$ it resembles that of a damped oscillation with a mean value equal to that of the transfer number for the solution presented by (Kronig and Brink, 1951). The transfer numbers for the solutions presented by (Newman, 1931) and (Kronig and Brink, 1951) are given as:

$$N_{t,nm} = 7.551 \qquad (29)$$
$$N_{t,kb} = 19.18 \qquad (30)$$

With these remarks in mind the following equations can represent the transfer numbers for the low Reynolds model:
For $Pe' \leq 1000$:

$$N_{t,low} = a\left[1 - b \cdot \exp\left(-c \cdot Pe'\right) \sin\left(d \cdot Pe' + e\right)\right] + N_{t,nm} \qquad (31)$$

For $Pe' > 1000$:

$$N_{t,low} = b \cdot \exp\left(-c \cdot Pe'\right) \sin\left(d \cdot Pe' - e\right) + N_{t,kb} \qquad (32)$$

The coefficients for Equations 31 and 32 are presented in Table 1 below:

Table 1. Coefficients for the low Reynolds model fit.

Model	a	b	c	d	e
Low	1.249×10^1	1.030×10^0	1.269×10^{-3}	5.169×10^{-3}	1.677×10^0
High	-	3.013×10^0	5.149×10^{-4}	6.508×10^{-3}	2.575×10^0

The maximum error between the fit for the low Reynold model and numerical results is 3.8% and the mean error is 0.63%, which are acceptable considering numerical uncertainties.

The fit for the high Reynolds model is simply a linear equation, which is given as:

$$N_{t,high} = 1.523 \times 10^{-2} \cdot Pe' \qquad (33)$$

The maximum and mean errors for the high Reynold model are 0.035% and 0.025%, respectively.

With the fits for the two models made these are combined into a single expression for the transfer number. It is proposed that the switching between the two models are determined by a switching Reynolds number, such that:

$$N_t = \text{neg}(Re - Re_{sw})N_{t,low} + \text{pos0}(Re - Re_{sw})N_{t,high} \qquad (34)$$

where Re is the external Reynolds number and Re_{sw} is the switching Reynolds number.
The function $\text{neg}(x) = 1$ for all negative values and $\text{neg}(x) = 0$ for all positive. The function pos0 is the opposite and includes 0 such that $\text{pos0}(0) = 1$.

5 Validation

For the model validation the experimental data from (Yao and Schrock, 1976) is used. In this study water droplets with diameters of 3, 4 ,5 and 6 mm are investigated . The average temperature of the droplets are sampled at varying points throughout a free fall of 3 m. Yao and Schrock (1976), made three experiments per droplet diameter at varying relative humidity of the surrounding air.

The experimental setup is modelled in an Euler-Euler framework, using the OpenFOAM solver reactingTwoPhaseEulerFoam (Rusche, 2003). The governing equations for the two-phase system are given as:

$$\frac{\partial}{\partial t}\left(\alpha_i \rho_i u_i\right) + \nabla \cdot \left(\alpha_i \rho_i u_i\right) = \sum_j \dot{m}_{ij} \qquad (35)$$

$$\frac{\partial}{\partial t}\left(\alpha_i \rho_i u_i\right) + \nabla \cdot \left(\alpha_i \rho_i u_i\right) = -\alpha_i \nabla p + \nabla \cdot \left(\alpha_i \mu_i \nabla u_i\right) + \sum_j \vec{F}_{ij} \qquad (36)$$

where α_i is the volume faction of the ith phase, ρ_i is the density of the ith phase, u_i is the ith phase velocity, $\sum_j \dot{m}$ is the mass transfer between the phases, p is the common pressure, and $\vec{F}_{ij}$ is the volumetric forces acting on the ith phase.

The forces considered to match the experimental contritions are gravity and drag. Where the drag coefficient for the liquid droplets are given as (Holterman, 2003):

$$C_d = \left[\left(\frac{24}{Re}\right)^{0.52} + 0.32^{0.52}\right]^{1/0.52} \qquad (37)$$

To account the effect of the relative humidity of the surrounding air, the temperature and mass concentration at the interface between the droplet and air has to be modelled. The temperature at the interface can be calculated by combining Equations 1 and 2 and adding the heat of evaporation. To solve this equation Newton acceleration is used. The initial guess is calculated without mass transfer. The interfacial temperature can be updated as follows:

$$T^{if} = T^{if} - \frac{\kappa_g(T^{if} - T_g) + \kappa_g(T^{if} - T_d) + \dot{m}}{\kappa_g + \kappa_d + \frac{d\dot{m}}{dT}} \qquad (38)$$

The mass transfer in the form of evaporation or condensation at the interface are determined by the saturation pressure of water vapour, which is given by the Arden Buck equation:

$$p_{sat}(T_c) =$$
$$611.21 \exp\left(\left(18.678 - \frac{T_c}{234.5}\right)\left(\frac{T_c}{257.14 + T_c}\right)\right) \quad (39)$$

where p_{sat} is the saturation pressure given in Pa and T_c is the temperature in degrees Celsius.

The Nusselts correlation for the external side of the droplet are given as (Ahmed and M.Yovanovich, 1994)

$$Nu = 2 + 0.775 Re^{0.5} \frac{Pr^{1/3}/\sqrt{2\gamma+1}}{\left[1 + \left(\frac{1}{(2\gamma+1)^3 Pr}\right)\right]^{1/6}} \quad (40)$$

$$\gamma = \frac{1}{(1+Re)^{0.25}}$$

The simulation is 2-D, were a small volume at the top of the domain is prescribed a volumetric fraction of water of 0.005 with the temperature reported by (Yao and Schrock, 1976). The small volume of water and the computational mesh is shown below:
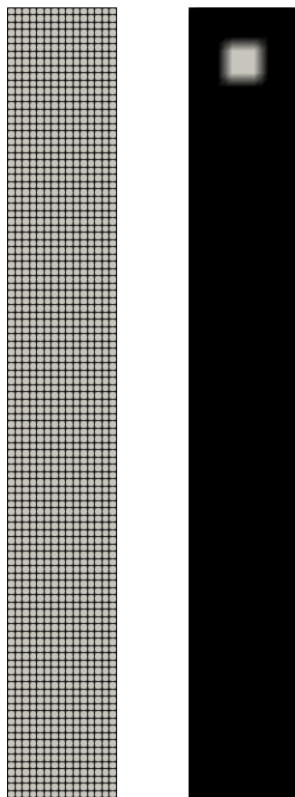
Figure 6. Computational mesh and the droplets at rest in the volume at the top of the domain.

In the simulation the sides are given a symmetric boundary condition, to avoid any wall effects. The total length of the domain is 4m to ensure that the boundary at the bottom does not interfere with the results. The temperature of the air is kept at a constant temperature, corresponding to that reported for the individual experiments.

It was found the switching Reynolds number Re_{sw} proposed in equation 34 should be 400, to produce the best match with the experimental results.

The results of the simulations are compared with the experimental results, and are shown in Figure 7:
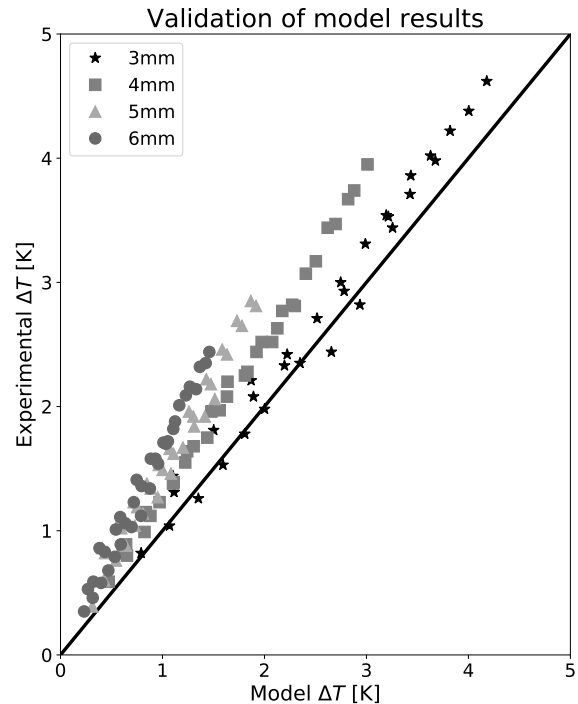
Figure 7. Simulations results compared with the experimental results of (Yao and Schrock, 1976). The results are presented as temperature decrease of the water droplets as it falls from rest.

It can be seen from Figure 7 that the simulation under predicts the decrease in water temperature as the diameter of the droplets increases. In the worst cases the results deviate by 40% compared to the experimental data.

6 Conclusion

Two models for large droplets with internal circulation have been investigated and an expression for the transfer number, Nusselt or Sherwood, has been found as a function of the modified Peclet number. The equations describing the fit of the numerical results were found to be within a satisfactory margin of error compared to the numerical results with a maximum error of 3.8%. A switching Reynolds number was proposed as a method for combining the two models.

The model for the droplets was implemented in a two-phase simulation, and compared to experimental data for

droplets experiencing free fall. The results of the simulations were compared with the experimental data and it was found that it under predicts the temperature decrease of the falling droplets. As the diameter of the droplet increases so does the error between the simulation and experimental data. In the worst case the simulated results deviates as much as 40%.

References

G. Ahmed and M.Yovanovich. Approximate analytical solution of forced convection heat transfer from isothermal spheres for all prandtl numbers. *Journal of Heat Transfer-transactions of The Asme - J HEAT TRANSFER*, 1994. doi:10.1115/1.2911456.

William Victor Andoe. *Liquid extraction with single droplets-effect of chemical reaction*. PhD thesis, Missouri University of Science and Technology, 1968.

R. Byron Bird, Warren E. Stewart, and Edwin N. Lightfoot. *Transport Phenomena*. John Wiley & Sons, Inc, 2nd edition, 2002. ISBN 0-471-41077-2.

J.S. Hadamard. Mouvement permanent lent d'une sphere liquide et visqueuse dans un liquide visqueux. *C. R. Acad. Sci.*, 152: 1735–1752, 1911.

A.E. Handlos and T. Baron. Mass and heat transfer from drops in liquid-liquid extraction. *AIChE*, 3:127–136, 1957. doi:10.1002/aic.690030121.

H.J. Holterman. *Kinetics and evaporation of water drops in air*. CIP-DATA ROYAL LIBRARY, DEN HAAG, NL, 2003. ISBN 90-5406-234-7.

R. Kronig and J.C. Brink. On the theory of extraction from falling droplets. *Applied Scientific Research*, (2), 1951. doi:10.1007/BF00411978.

Bruce R. Munson, Ted H. Okiishi, Wade W. Huebsch, and Alric P. Rothmayer. *Fluid Mechanics*. John Wiley & Sons, 7th edition, 2013. ISBN 978-1-118-318676.

Albert Broadus Newman. The drying of porous solid: Diffusions and surface emission equations. *AIChE*, 27:203–220, 1931.

Donald R. Olander. The handlos-baron drop extraction model. *AIChE Journal*, 12(5):1018–1019, 1966. doi:10.1002/aic.690120533.

Douglas L. Oliver and Adham W. Souccar. Heat transfer from a translating droplet at high peclet numbers: Revisiting the classic solution of kronig & brink. *Journal of Heat Transfer*, 128(7):648–652, 2006. doi:10.1115/1.2193542.

Henrik Rusche. *Computational fluid dynamics of dispersed two-phase flows at high phase fractions*. PhD thesis, Imperial College London, January 2003.

R. M. Wellek, W. V. Andoe, and R. J. Brunson. Mass transfer with chemical reaction inside single droplets and gas bubbles: Mathematical mechanisms. *Can. J. Chem. Eng*, 48(6):645–655, 1970. doi:10.1002/cjce.5450480601.

H. G. Weller, G. Tabor, H. Jasak, and C. Fureby. A tensorial approach to computational continuum mechanics using object-oriented techniques. *Computers in Physics*, 12(6):620, 1998. doi:10.1063/1.168744.

Shi-Chune Yao and V. E. Schrock. Heat and mass transfer from freely falling drops. *Journal of Heat Transfer*, 98(1):120–126, February 1976. ISSN 0022-1481. doi:10.1115/1.3450453.

Nomenclature

Symbol	Description	Unit
a	Droplet radius	m
A_n	Coefficient for Equation 15	-
C_d	Drag coefficient	-
$\vec{F}$	Volumetric force	N m^{-3}
j	Heat or mass flux	J m^{-2} s^{-1} kg m^{-2} s^{-1}
J	Dimensionless surface flux	-
$\dot{m}$	Mass flow	kg s^{-1}
N_t	Transfer number	-
Nu	Nusselts number	-
p	Pressure	Pa
Pe	Peclet number	-
Pe'	Modified Peclet number	-
Pr	Prandlt number	-
r	Radial coordinate	m
R	Dimensionless radial coordinate	-
Re	Reynold number	-
Sh	Sherwood Number	-
t	Time	s
T	Temperature	K
u	Velocity	m s^{-1}
V	Volume	m^3
α	Volume fraction	-
Γ	Thermal conductivity or mass diffusivity	m^2 s^{-1}
ζ	Dimensionless tori centred radial coordinate	-
θ	Angle	-
κ	Heat or mass transfer coefficient	W m^{-2} K^{-1} m s^{-1}
λ_n	Eigenvalue for Equation 15	-
μ	Viscosity	Pa s^{-1}
ρ	Density	kg m^{-3}
τ	Dimensionless time/ Fourier number	-
Φ	Dimensionless potential	-
$\bar{\Phi}$	Volume averaged dimensionless potential	-
Ψ	Streamline	m^3 s^{-1}

Symbol	Description
subscripts	
c	Celsius
d	Droplet
g	Gas
$high$	High Reynolds model
i	Phase i
j	Phase j
low	Low Reynolds model
nm	Newman solution
kb	Kronig and Brink solution
rel	Relative
r	Radial
sw	Switching
θ	Tangential
$+1$	Step forward
-1	Step backwards
superscripts	
if	Interface
abbreviations	
CFD	Computational Fluid Dynamics

Improving gas distribution and evaporation in a gas cooling tower at Norcem Brevik

Fredrik Kasin*, Are Larsen, Christian Andersen, Rajan Jaiswal, Rajan K. Thapa

Department of Process, Energy and Environmental Technology, University of South-Eastern Norway,
`fkasin@hotmail.com`, `arelarsen@hotmail.com`, `christiana.1990@gmail.com`
`{rajan.jaiswal,rajan.k.thapa}@usn.no`

Abstract

In this work the effect of the geometry of a GCT (Gas Cooling Tower) on the flue gas distribution in a cooling tower in Norcem, Brevik is investigated by using CPFD (Computational Particle Fluid Dynamics) modeling and simulation. Barracuda Virtual Reactor is used to perform the simulations of a baseline model, representative of the existing cooling tower. Results from simulations show poor distribution of flue gas and recirculation zones occurring on both sides of the distribution screen. The cause of this uneven distribution is attributed to challenging geometry and poor screen performance. To counteract this problem, a new model is developed with a second screen and guide vanes. The screens are placed in the lower part of the GCT's diffusor, whilst the guide vanes are in the diffusor's inlet duct. The screens and guide vanes are modeled as baffle computational cells with zero thickness. The implementation of a second screen eliminates the recirculation zone below the screens and improve the distribution.

Keywords: Norcem, Brevik, CPFD, cooling tower, gas distribution, Barracuda

1 Introduction

Concrete is the most used building material in the world. Norcem AS annually produces 1.7 million tons of cement at its two plants in Norway. The Brevik plant has always focused on technological innovation and has therefore been at the frontline of efficient and environmentally friendly cement production. Cement production is an energy-intensive process and reuse of waste heat is an important part of this. Waste heat from the preheater is used for drying raw materials, and the temperature must, therefore, be controlled. The temperature of the preheater flue gas is about 380-410°C. The gas should have a temperature of about 190-290°C to use for drying of the raw materials and prevent damage to equipment. To control the temperature, the gas is cooled in a cooling tower by water spray. The cooling tower is designed as a cylinder, with a maximum diameter of 6 m. Figure 1 shows the upper part of the cooling tower and a gas distribution screen. The gas is fed into the top of the tower and passes through a diffuser. In the lower part of the diffuser there is a distribution screen which has a thickness of 3mm and 70mm holes throughout the cross section. The aim of the screen is to distribute gas evenly throughout the cross-

sectional area of the tower. The gas is cooled by cold water injected through 16 nozzles placed evenly around the perimeter of the tower at 1600mm below the distribution screen. Each nozzle is placed 22.5° apart from each other and they have pressure 25-27 bar.
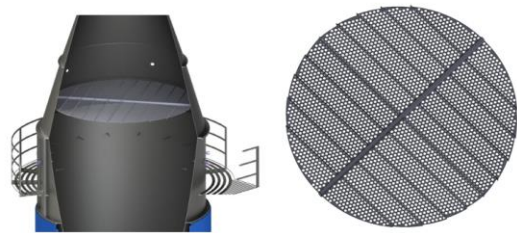
Figure 1. Upper part of cooling tower and gas distribution screen

Today's design makes use of only one gas distribution screen which results in insufficient distribution thus incomplete evaporation of cooling water. This creates spilling formation at the bottom of the cooling tower. This prevents deposition of dust and results in build ups and increased wear which leads to reduced production capacity. It is suspected that the screen is not able to distribute the gas evenly. Therefore, it is important to understand the actual conditions that causes the challenges and look at possible measures.

To investigate the causes that create the problems, data for geometry and flow parameters along the cooling tower are obtained from Norcem and analyzed to make a simulation model in CPFD software Barracuda VR. This will provide a basis for how the gas distribution can be improved. When using simulation tools, consideration is given to what kind of changes should be made and what effects this will have.

2 Computational model

In this work, a Computational Particle Fluid Dynamics (CPFD) model is used to calculate gas, particle, and water flow along the cooling tower. Gas and particles are passing through the distribution screen. All simulations are run in Barracuda Virtual Reactor 17. Barracuda VR uses the Multi-Phase Particle in Cell (MP-PIC) approach, where combined Eulerian and Lagrangian methods are used for the modeling of fluid-solid interaction. The fluid phase is solved as the continuum, while the particle phase is solved using the Lagrangian method. Langrangian method follows a

particle independent of the grid, and its properties are calculated at each point/time step. In Eulerian, control volume method, the geometry is divided into several control volumes, and what goes in must go out. The number of control volumes are determined by the size of the grid used. The equations are solved at each point of the grid (Falkovich, 2011). The equations involved in the computational model presented in this section are proposed by O'Rourke *et al*. and Snider *et al*. and the detail about the governing equation and computational schemes can be found there (Andrews, 1996; Sinder, 2001).

The fluid phase is described with Navier-Stokes mass, impulse, and energy equation (Thapa, 2015). They describe how the pressure, velocity, temperature, viscosity, and density of a moving gas are related. The equations contain several dependent variables, such as velocity and viscosity, as a function of four independent variables x, y, z, and t. The equations are thus partial differential equations (Hall, 2015)

One of the most important parameters in CPFD is the drag model. The drag model indicates forces acting on a particle from a fluid. In simulations, it is chosen to use Gidaspow's model for both particles and water. This model is based on Wen & Yu and Ergun's models. Wen & Yu's model is adapted to flows with low particle fractions. Ergun's model is adapted to large volume fractions and will not be used in this work.

2.1 Model parameter

The composition of the flue gas coming into the cooling tower consists of several fractions, but the main constituents are nitrogen, carbon dioxide, oxygen, water vapor, and carbon monoxide. The volume fraction of these will vary with the type of clinker produced and how the process is operated. What types of fuel and quantities of these will be important factors. The composition used in this report is from the production of Standard clinker and measured in connection with a heat balance performed in 2018 and is shown in Table .

Table 1. Composition of flue gas

Component	Volume fraction	Molecular weight[g/mol]
N_2	62.7%	28.0134
CO_2	23.5%	44.0095
H_2O	9%	18.0152
O_2	4.6%	31.9988
CO	2%	28.0101

The models include data on the particles that are part of the simulation. This includes particle type, size distribution, sphericity, and particle density. Raw meal has been used as a starting point for particle parameters,

although this will be a mixture of raw meal, calcined raw meal, and ash residue. The size distribution for solid particles is an average distribution over a year of production of Standard raw meal. The particle size distribution (PSD) used in the computational model is shown in Figure 2.
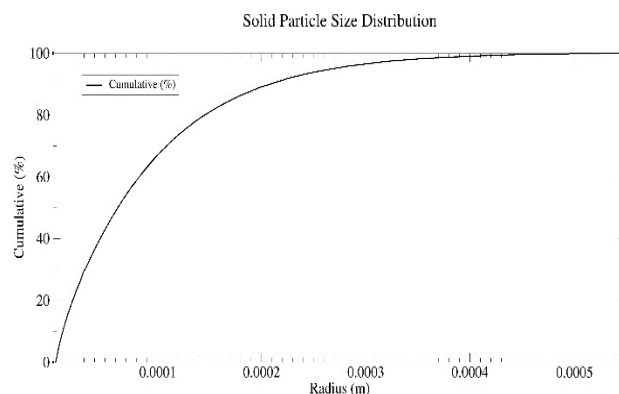
Figure 2. Solid particle size distribution, for raw meal

For the initial simulations only water droplets are added to the model. The droplet size distribution is shown in Figure 3.
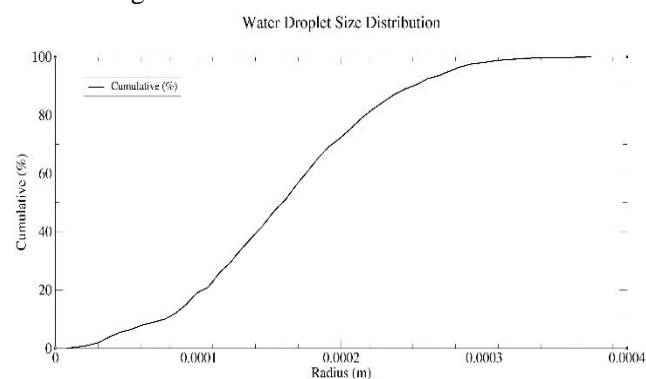
Figure 3. Water droplet size distribution

The droplet size effects the overall performance of the cooling tower. Fine droplets accelerate the evaporation process and hiders transportation of moisture out of the tower.

2.2 Model geometry and boundary conditions

A model of geometry is prepared for the cooling tower. To model the cooling tower on the correct scale and with the necessary level of detail, production drawings of the inlet duct, cooling tower, and outlet provided by Norcem have been used. The drawings are hand drawn from the period 1978 to 1986. The drawings are the starting point for the model. Some of the parameters were missing and new physical measurements have been made to find out the missing geometry parameters. This applies mainly to the lower part of the tower with piles and outlet duct. New measurements have yielded a margin of error against original drawings by 0.07%.

CPFD software Barracuda reads only massive parts of the model as a flow path. The tower is thus modeled as a massive model where nozzle pipes are modeled as cavities. The boundary to the surroundings is then determined by the inner diameter of the tower mantle.

The model is drawn according to the right-hand rule so that the z-axis is the vertical axis in the model. To make sure this is correct, the model is drawn in a positive z-axis. This is done so that the axis representing the gravitational direction should correspond between the two programs. The only simplification of the model that has been done is to omit the screw conveyor at the bottom as it will not significantly affect the flow. The geometry was drawn in CAD software SolidWorks, saved as an STL file, and then imported to Barracuda. The model geometry of the tower and boundary conditions are shown in Figure 4..
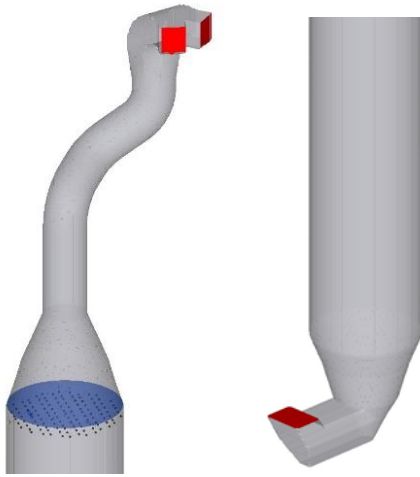
Figure 4. Model geometry and boundary conditions

For the isothermal gas flow simulations both inlet and outlet of the cooling tower are defined as flow boundary. This is because the inflow and outflow are set equal. When injection of water is added to the simulation the outlet boundary is changed to pressure BC. This defines outlet conditions based on pressure. Injection boundaries are used for water nozzles.

3   Simulation methods

A series of simulations were run using the baseline model. Before simulations, physical monitors were installed under and above the distribution screen and the pressure drop was measured to 70Pa. This measurement is used in the CPFD model.

The main objective of this work is to improve the gas distribution so that better evaporation of the cooling water can be achieved. To measure an improvement, a list of criteria has been set for velocity differences over the tower cross-section:

- Measurements must be made across the entire cross-section, except for a peripheral area behind

nozzles. This area should not exceed 10% of the total cross-section.

- Horizontal distance between the measuring points should not exceed 500mm.

- Recirculation zones must not occur within the evaporation zone.

- The velocity deviations in the measuring points should not exceed +/- 30% of the calculated velocity for a given cross-section. This gives a permissible deviation from -2.26 to -4.12m / s.

- Speed deviations in the peripheral zone must not exceed +/- 50% of the calculated speed.

- The standard deviation for all measurement points should not exceed 18%.

4   Results and discussion

4.1   Baseline model with one distribution screen

The first simulations were run for the entire tower without dust particle loading and water injection. The objective was to get an overview on how the gas flows in the area above and under the distribution screen where water nozzles are located. It is assumed that the volume fraction of dust is small enough that it does not significantly affect the gas flow. The simulation is set up with inlet and outlet Flow BC with equal mass flow and temperature. The boundary conditions for this model are shown in Table 2.

Table 2. Flow boundary conditions

	Pressure [Pa]	Temperature [K]	Mass flow [kg/s]
Inlet BC	92 073	673.15	52.15
Outlet BC	91 750	673.15	-52.15

In the cooling tower, the temperature variation, during normal operating conditions, does not exceed 1-2 degree Celsius. Therefore, the temperature at inlet and outlet boundaries are considered as constant.

The geometry of the model is divided into one million uniform cells as shown in Figure 5. which gives a cell size of 17cm. The grid size was not possible to reduce further because of very large computational time. If the grid were reduced further, it would have taken few weeks for one simulation which is the beyond of the scope of this work.
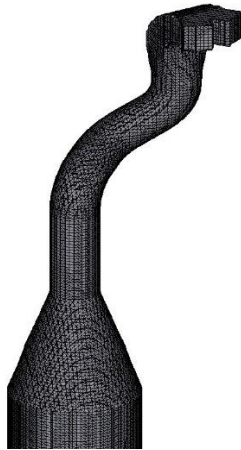
Figure 5. Grid of the geometry

The distribution screen is modeled similarly to a porous plate with a calculated resistance. In such a plate, particles will not be able to pass, and we only get a picture on how the gas flows. This is done as a simplification to save simulation time. For the current model, the simulation time is 12 hours. A modeling of the distribution screen with 3 mm thickness would have increased the need for cells to several billion and given a simulation time that is far from the possibilities of this work.

Results from this model show large velocity variations over the tower cross-section. Figure 6. shows the axial velocity in the XY plane at the level of water nozzles and the central axial velocity in the YZ plane.
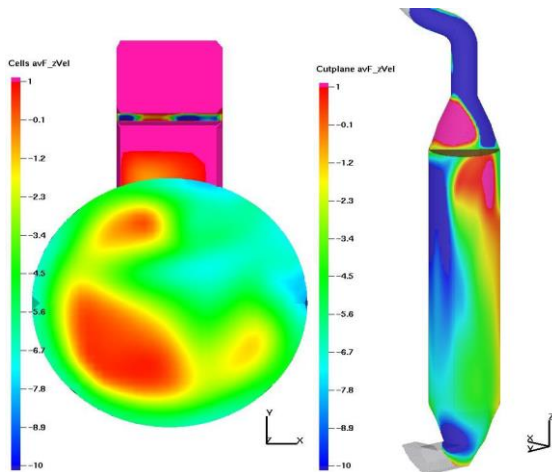
Figure 6. Velocity profile for the baseline model, one screen

From Figure 6., a clear distortion is seen under the distribution screen, where the gas flow changes the direction. The radial velocities increase in the negative x-direction up to the wall of the diffuser. Here the direction of flow changes and the axial velocity becomes positive. This is due to the geometry of the duct from the upper cyclone stage down to the diffuser. After the last duct bend, wall separation occurs, and the gas is pressed against one side, which continues all the way down to the distribution screen. Here, a recirculation zone is

formed which further amplifies the effect of the skewed gas distribution. Consequently, the gas velocity through the screen becomes uneven.

As a consequence, a large recirculation area and stagnant gas pockets under the distribution screen. In these areas, evaporation of cooling water will be greatly reduced. The radial velocity components along the x and y axes exceed the axial velocity in several areas. This will transport cooling water and particles to the tower walls where they agglomerate (Nielsen, 2002).

The simulation results from baseline model with one distribution screen show the following:

- The standard deviation of the velocity profile at the level of the water jets is 65.13%
- For single measurements, the velocity ranges from 0.5 to -7.05m / s
- The average of each measurement point shows 50 measurements where the velocity is higher than -4.21m / s
- 34 measuring point has an average value where the velocity is lower than -2.26m / s
- A total of 84 of 119 measurements are outside the recommended range.

The pressure loss from the tower inlet to below the screen is 269Pa.

4.2 Model with two distribution screens

In the next model a second distribution screen has been added 1.95m above the original one. The results from this simulation show a significant improvement of the gas distribution. Recirculation zones and stagnant gas pockets are eliminated, and the radial velocities are greatly reduced. Figure 7. shows the axial velocity in the XY plane at the level of water nozzles and the central axial velocity in the YZ plane.
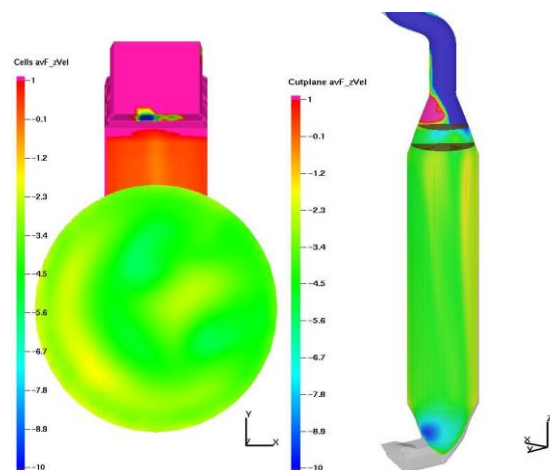
Figure 7. Velocity profile for model with two screens

The major results from the second model summarized:

- The standard deviation of the velocity profile at the level of water jets is 21%
- The velocity ranges from -1.75 to -5.52m/s

A total of 43 of 119 measurements are
outside the recommended range.

The pressure loss for this model is 278Pa, an increase
of 9Pa from the baseline model.

4.3 Models with water injection
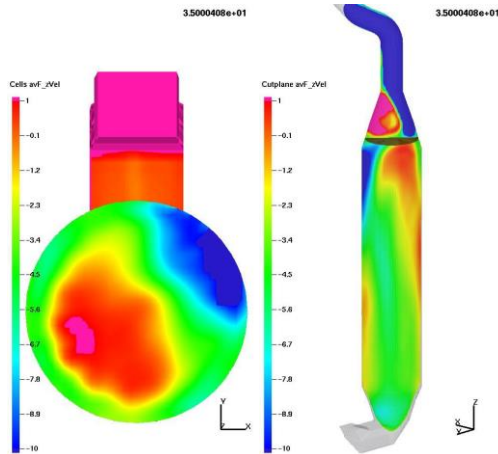
Figure 8. Velocity profile with water injection, one screen

Based on the baseline and second model, water
injection was added to see how this affects the gas
distribution and pressure loss. The model is set up with
inlet Flow BC and outlet Pressure BC. Outlet pressure
is specified to represent the measured value. Water is
injected through an Injection BC via 16 simplex
spillback atomizers placed 500mm below the diffuser at
the rate of 3.74kg/s. For the baseline model an increased
velocity deviation and strengthening of recirculation
zone is seen (see Figure 6.). The recirculation zone
creates a zone of stagnant gas along the tower wall
where water spray and particles will agglomerate and
cause corrosion. This area correlates with tower wall
corrosion seen over the past few years. Opposite to the
recirculation zone, a high velocity zone is created where
complete evaporation is not achieved within the tower.
The pressure loss from inlet to below screen is 858Pa.
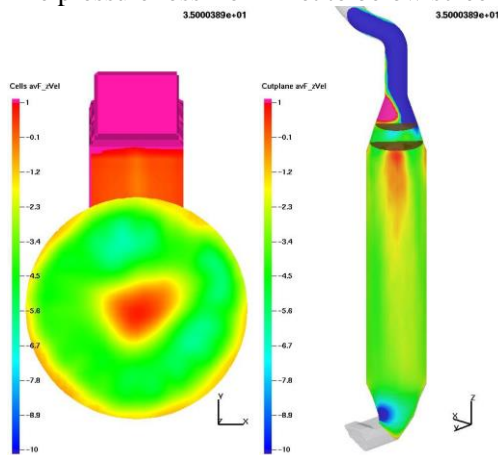
Figure 9. Velocity profile with water injection, two screens

For the model with two screens, see Figure 9. a
recirculation zone is created centrally in the tower. The
following inward radial velocities prevents the

transportation of particles to the tower walls. This model
shows almost complete evaporation of water droplets
before flue gas exits tower. By using flux planes this
amount is calculated to 6.6kg/min at tower exit. The
pressure loss from inlet to below screens is 859Pa. Thus,
the impact of a second screen on gas transportation is
negligible.

The results show that the main contributor to pressure
loss is water injection and not the distribution screen(s).
This occurs despite a lower outlet gas velocity.

4.4 Models with new nozzles

Simulations were further run for the tower with new
nozzle inclined at 45° with spray angle of 60°. The
droplet size distributions are kept similar to the prior
simulations. All other parameters are also kept
unchanged. The first simulation were run with one
distribution screen. The results show more water and
particle accumulation near the wall of the tower which
will negatively effect on evaporation process and gas
distribution.
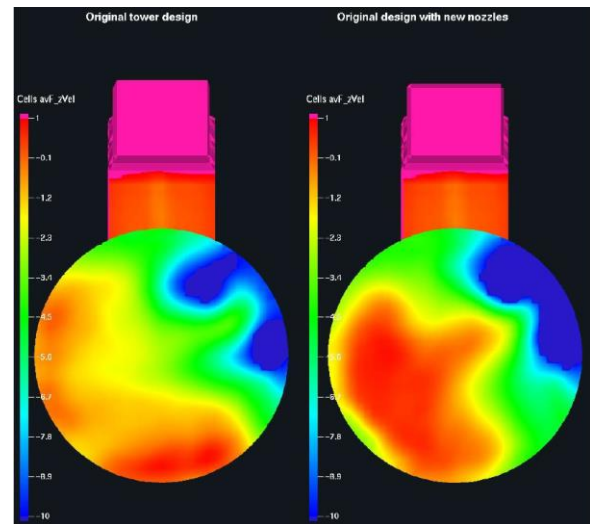
Figure 10. Comparison of gas velocities at 2500 mm below
the distribution plate

Simulations were also run for the tower with two
distribution screens and new nozzles. The gas velocity
distribution along the tower is presented in Figure 11.
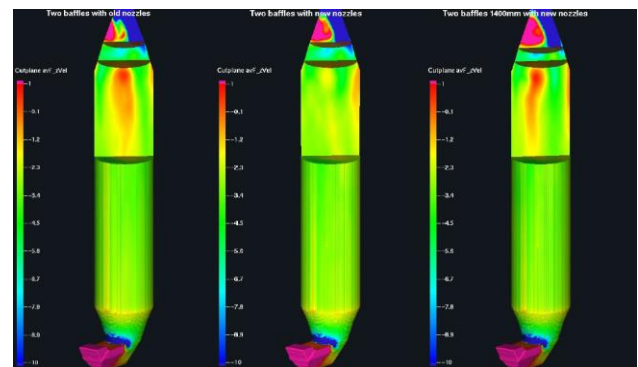
Figure 11. Comparison of gas velocities along the center
of the tower

Figure shows that the gas distribution at the tower with new nozzles and two distribution screens is improved.

5   Conclusions

Results from simulations done on the second model with two distribution screens show an improved gas distribution and complete evaporation of cooling water. This is achieved without significant increase in the pressure loss through the tower. Nevertheless, some recirculation occurs along the walls in the upper part of the evaporation zone. This is due to wall separation below the lowermost distribution screen. Small radial velocities that occur may transport particles to tower walls in this area. The gas distribution in the tower with two distribution screen and new nozzles gives better results.

References

M. J. Andrews and P. J. O'Rourke. The multiphase particle-in-cell (MP-PIC) method for dense particulate flows, *International Journal of Multiphase Flow*, 22(2): 379-402, 1996.

G.   Falkovich. *Fluid Mechanics: A Short Course for Physicists.* Cambridge University Press, 2011.

N.   Hall.   *Navier-Stokes Equations*.   NASA,   05.2015 https://www.grc.nasa.gov/WWW/k12/airplane/nseqs.html

N. F. Nielsen and L. Lind. Applying CFD for Design of Gas Conditioning Towers with Swirling Flow, presented at the *ASME 2002 Pressure Vessels and Piping Conference, Vancouver*, 2002. https://doi.org/10.1115/PVP2002-1588

D.M.   Snider,   P.J.   O'Rourke,   and   M.J.   Andrews.   An incompressible three-dimensional multiphase particle-in-cell   model   for   dense   particle   flows,   *Journal of Computational   Physics*,   170(2):   523-549,   2001.

R. K. Thapa, A. Frohner, G. Tondl, C. Pfeifer, and B. M. Halvorsen. Circulating fluidized bed combustion reactor: Computational Particle Fluid Dynamic model validation and   gas   feed   position   optimization,   2015. DOI: https://www.sciencedirect.com/science/article/pii/S0 098135416301661.

A new Approach for Scaling Up from a Small Cold to a Large Hot Bed for Biomass Conversion in a Bubbling Fluidized Bed Reactor

Cornelius E. Agu[1], Rajan K. Thapa[1], Christoph Pfeifer[2], Lars-Andre Tokheim[1], Britt M.E. Moldestad[1]

[1]Department of Process, Energy and Environmental Technology, University of South-Eastern Norway, 3918 Porsgrunn, Norway, {Cornelius.e.agu, rajan.k.thapa, Lars.A.Tokheim, britt.moldestad}@usn.no
[1]Department of Material Sciences and Process Engineering, University of Natural Resources and Life Sciences, 1190 Vienna, Austria, christoph.pfeifer@boku.ac.at

Abstract

Bubbling fluidized beds are simple and attractive means of achieving efficient conversion of biomass if particle segregation and the associated effects are minimized. To improve the knowledge of fluidized bed reactor design, this paper compares the behavior of a hot bed containing a certain amount of biomass with the behavior in a cold bed having the same biomass loads and particle properties. An approach for scaling up a cold bed to a large hot bed for the same volume fraction of biomass is introduced. The proposed scheme uses the bed expansion ratio as an output from the cold bed. This approach provides an accurate means of attaining dynamic similarity in bubbling behavior between two different beds without constraining the fluid and particle properties as well as the bed height.

Keywords: biomass, scale up, bed expansion, CPFD, bubbling fluidized bed

1 Introduction

To enhance the understanding of fluidized bed behavior for reactor designs and scale up, this paper compares the bubbling behavior in a hot fluidized bed containing biomass with that observed in Agu et al. (2019a) under cold flow condition using the same bed material and biomass properties. In the literature (Agu et al., 2019a), the mixing and segregation pattern of wood chips and pellets of a wide density difference but similar volume-equivalent spherical particle diameter was studied using electrical capacitance tomography (ECT).

Understanding how the cold bed behavior is related to the behavior in an active hot bed is often a challenge in modelling and scaling up. The traditional scaling up approaches (Glicksman et al., 1993; Glicksman, 1988; Glicksman, 1984; Romero and Johanson, 1962) that have been used in many studies (Kraft et al., 2018; Biglari et al., 2016; Thapa and Halvorsen, 2014) do not consider the influence of feedstock and volatiles (i.e. biomass) on the bubbling behavior. As shown in different studies (Agu et al, 2019a; Fotovat et al, 2015a), the amount of biomass in a bed affects the bubble diameter, bubble frequency and bed expansion, and also causes a delay in slug flows. In an attempt to keep the ratios of diameter and density of fuel particles to those of a bed material constant, the two beds will have different fluid and particles properties. Different scaling up approaches accounting for intrinsic behavior in a bed have also been demonstrated. On the account of segregation effects, a scaling approach based on the biomass residence time distribution was reported in (Zhang and Xu, 2015). Scaling up a bed based on the segregation effects on the thermochemical conversion of biomass can also be achieved by using the relative time scales (Damköhler numbers) between the segregation and devolatilization of the fuel particles (Salatino and Solimene, 2017). Moreover, a bubbling fluidized bed can also be scaled up by keeping at least one of the bubble properties constant (Agu et al, 2019b).

In this study, a new approach for scaling a bubbling bed behavior from a cold small to a hot large bed is introduced based on a combination of Horio et al (1986) scaling law and correlations for minimum fluidization velocity and bed expansion ratio. The scaling approach introduced allows the same properties of biomass as well as the same bed material type to be used in both beds. For the scaled bed, the Wen and Yu (1966a) correlation is used to determine the particle size while the model for bed expansion ratio (Agu et al., 2019c) is used to determine the aspect ratio (bed height to diameter) owing to the fact that bubble growth and bed expansion depend on the bed aspect ratio. The scaled bed aspect ratio is also verified by simulation of the two different beds using the Barracuda software based on the CPFD (computational particle-fluid dynamics) model.

2 Modelling and scaling up

To scale up fluidized bed behavior from a small to a larger bed diameter, scaled and transferrable information from the smaller (reference) bed is required. In a hot flow bed, temperature and pressure are easier to measure. Temperature is not a product of bubble flows but the distribution of temperature as well as the pressure fluctuations can be linked to the bubbling behavior. However, modelling of temperature distribution and pressure fluctuation is a complex task.

In a cold bed, different properties can easily be measured, including bubble diameter, bubble frequency and bed expansion. Different correlations are also available in literature for prediction of the different bubble properties for a bed of single particle type, e.g. sand particles. Based on the results reported in Agu et al (2019a) for binary mixtures of sand and biomass, none of the available models (Choi et al, 1988; Darton et al, 1977; Mori and Wen, 1975) for bubble diameter can predict the bed behavior with a good accuracy. The bed expansion predicted using different models (Agu et al, 2019c; Hepbasli, 1998; Werther, 1978) at $U_0/U_{mfs} = 2$ show similar trends as the experimental data. Numerically, the Agu et al (2019c) model agrees very well with the data from the pellet beds while for the wood chip beds, a good agreement was obtained when the segregated layer at the surface of the bed is excluded in the total bed height measured. The bed expansion model as proposed by Agu et al (2019c) is described below:

$$\Delta e = \left[1 - 0.103\left(U_0 - U_{mf}\right)^{-0.362}\left(\frac{\bar{d}_b}{D}\right)\right]^{-1} - 1 \quad (1)$$

$$\bar{d}_b/D = 0.848\left(\frac{U_0}{D}\right)^{0.66}\left(1 - c\left(\frac{U_0}{U_{mf}}\right)^{a-1}\right)^{0.66} ;$$

$$Ar > 400 \text{ and } U_0/U_{mf} > c^{1/(1-a)} \quad (2)$$

where $\Delta e = (H_f - H_{mf})/H_{mf}$ is the bed expansion and $\bar{d}_b/D$ is the bubble to bed diameter ratio averaged over the bed height. H_{mf} and H_f are bed heights at the mixture minimum fluidization velocity U_{mf} and at the operating gas velocity U_0, respectively. Eq. (1) is semi-empirical and was developed based on the mass balance in the solid and gas phases. By using the particle-dependent average bubble diameter shown in Eq. (2) as proposed by Agu et al (2018), Eq. (1) predicts the behavior in beds with different particle properties. Ar in Eq. (2) is the Archimedes number, and the equation is applicable for large particles where $Ar > 400$ and for gas velocities U_0 such that $U_0/U_{mf} > c^{1/(1-a)}$ which corresponds to the minimum bubbling condition of the bed, where a and c can be obtained as described in the literature (Agu et al, 2018).

For a binary mixture of sand and biomass particles, the mean particle diameter d_m, density ρ_m and sphericity φ_m used for evaluation of $\bar{d}_b/D$ can be obtained from the following, where the subscripts "m", "b" and "s" denote mixture, biomass and sand, respectively, and y_b is the biomass volume fraction in the binary mixture:

$$d_m = \left[\frac{y_b}{\varphi_b d_b} + \frac{(1-y_b)}{\varphi_s d_s}\right]^{-1} \quad (3)$$

$$d_{p,sph,m} = \left[\frac{y_b}{d_b} + \frac{(1-y_b)}{d_s}\right]^{-1} \quad (4)$$

$$\rho_m = y_b\rho_b + (1-y_b)\rho_s \quad (5)$$

$$\varphi_m = d_m/d_{p,sph,m} \quad (6)$$

3 Results and discussion

Figure 1 compares the predictions of the bed expansion using Eq. (1) with the experimental data obtained at different gas velocities and biomass loads. For the bed of pellets, the mean bed particle properties are based on the total amount of biomass charged in the bed. However, in the wood chip bed, the computation of the mean particle properties is based on 50% of the total biomass charged since approximately 50% of the biomass load lies above the bed surface, preventing the bubble passage (Agu et al, 2019a). As can be seen in Figure 1(a), the model results agree very well with the measured data from the wood chip beds, particular with $y_b = 0.2$. For the pellet beds, the model prediction accuracy increases with increasing gas velocity. The inaccuracy in predicting the bed expansion can be associated with the strong interactions between the biomass and sand particles at lower gas velocities. With low gas velocity, a higher amount of sand particles are in the interstices between the biomass particles, thus reducing the bed voidage and expansion. Increasing the gas velocities increases the inter-particle spacing, thereby enhancing the bed expansion.

3.1 Scale up from cold small to hot large bed

To attain a dynamic similarity in the flow of bubbles between a cold small bed and a hot larger bed based on the same biomass properties and volume fraction, the Horio et al. (1986) scaling law as described below can be applied.

$$\frac{U_0 - U_{mf}}{\sqrt{gD}}, \frac{U_{mf}}{\sqrt{gD}} \quad (7)$$

The Horio et al (1986) law does not impose a restriction on the bed height and on the solid to fluid density ratio. In addition, Glicksman (1988) showed that (7) is equivalent to the scaling law proposed for the viscous limit set for two different beds of the same geometry and negligible solid to fluid density ratio. Implementation of Eq. (7) in this study for scaling up the biomass-sand bed behavior without considering the height and density ratios is based on the fact that bubble growth in a fluidized bed depends on the particle size, fluid properties and bed aspect ratio (Agu et al, 2019b). For larger particles or less viscous flow, the bubble growth rate with an increase in gas velocity is high. The bubble size increases with an increase in the bed aspect ratio, particularly for small particle sizes. From (7),
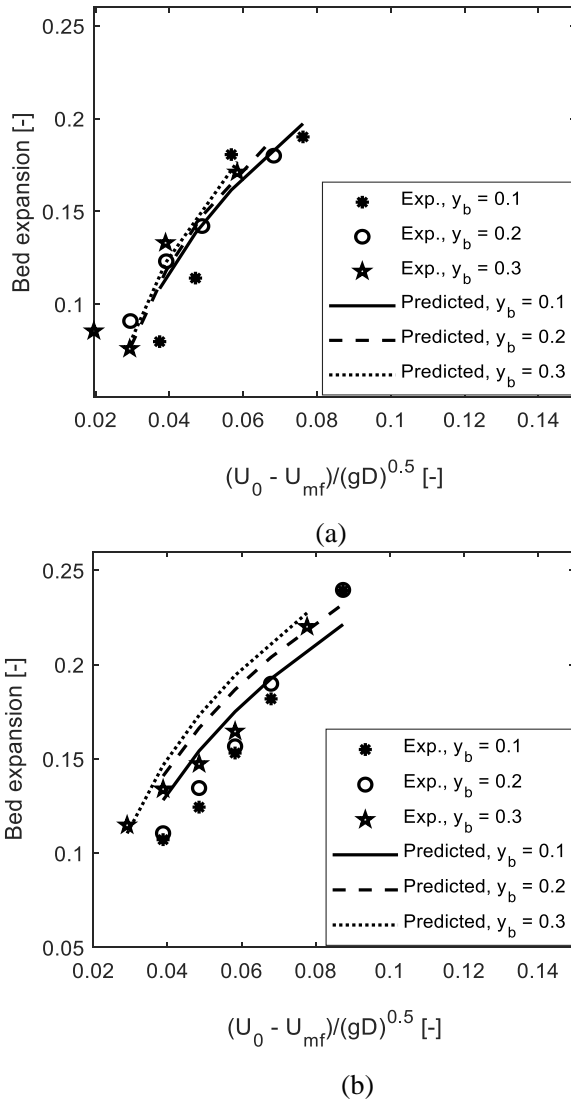
Figure 1. Bed expansion ratio, comparing the predicted results based on Eq. (1) with the experimental data obtained in the cold bed of diameter 10.4 cm at different biomass loads and air velocities as described in Agu et al. (2019a); (a) wood chips (b) wood pellets.

$$U_{mf2} = \sqrt{r}\, U_{mf1} \qquad (8)$$

where $r = D_2/D_1$ is the size ratio between the large bed diameter D_2 and the small bed diameter D_1, and U_{mf2} and U_{mf1} are the corresponding minimum fluidization velocity of the particles in the respective beds. Eq. (8) indicates that particles in the larger bed will have a larger particle size since $U_{mf2} > U_{mf1}$. The particle size for a given value of U_{mf} can be obtained from Eq. (9) as proposed by Wen and Yu (1966a).

$$\frac{\rho_g U_{mf} d_s}{\mu_g} = \left[-33.67 + \sqrt{((33.67)^2 + 0.0408 Ar)}\right];$$
$$Ar = \frac{d_s^3 \rho_g (\rho_s - \rho_g) g}{\mu_g^2} \qquad (9)$$

If the aspect ratio h_0/D is the same for both beds, the resulting bed height for the larger bed may be too high for industrial application; in addition, the bubble size at

the same $(U_0 - U_{mf})/\sqrt{gD}$ value may be larger. Therefore, for successful application of the Horio et al (1986) scaling law, an output from the large bed that must be correctly matched with that from the smaller bed is required. For the binary mixture of biomass and sand particles, the bed expansion ratio is the output applied in the scaling approach introduced in this study. The bed expansion of the scaled hot larger bed can be simulated from Eq. (1). By comparing the simulated result with the measured data from the cold small bed at the same biomass volume fraction, the appropriate bed aspect ratio can be determined. Since Eq. (1) accurately predicts the behavior in the cold bed of diameter 10.4 cm containing 20% wood chips, this bed is used as the reference here.

Figure 2 compares the bed expansion predicted for the scaled hot large bed with those measured in the cold small bed at different biomass loads. The new bed aspect ratio is evaluated as $h_{0*}/D_* = 2.7$, where $h_{0*} = 50$ cm is the initial height of the reference bed (cold bed) and $D_* = 18.5$ cm is the characteristic bed diameter obtained by fitting (1) at the hot bed condition with the data obtained from the cold bed at $y_b = 0.2$. As can be seen, there is a good agreement between the simulated results and the experimental values from the bed with the 20 vol.% wood chips. The similarity also observed for other wood chip beds and those of pellets confirms the validity of the scaling approach.

Different dimensionless groups associated with the two beds based on the bed materials are compared in Table 1. In addition to the Froude numbers, $U_{mf}^2/(gD)$, the particle Reynolds numbers, $\rho_g U_{mf} d_s/\mu_g$ and Archimedes numbers, $Ar = d_s^3 \rho_g (\rho_s - \rho_g) g/\mu_g^2$ are closely matched. To achieve the same h_0/d_s ratio, the aspect ratio of the larger bed has to be reduced to $h_0/D = 1.47$, which is possible if only the pure bed material is in the bed. However, by considering that the bed expansion ratio decreases with an amount of biomass in the bed, the height of the large bed needs to be as high as possible to achieve a similar behavior to that in the smaller bed diameter.

3.2 Verifying the scaled bed aspect ratio

To verify the aspect ratio of the scaled large bed, the two different beds (cold at ambient and hot at 800 ˚C), each containing 20 vol.% wood chips were simulated using Barracuda VR software. Barracuda is a commercially developed platform for implementing the computational particle-fluid dynamics (CPFD) scheme. CPFD is based on the multiphase-particle-in-cell (MP-PIC) concept introduced by Andrew and O'Rourke (1996). In the CPFD scheme, the Euler-Lagrangian modelling approach is applied for fluid volume and particle tracking in gas-solid systems. With the MP-PIC concept, the computational particle represents a large number of particles with similar properties.

(a)

(b)
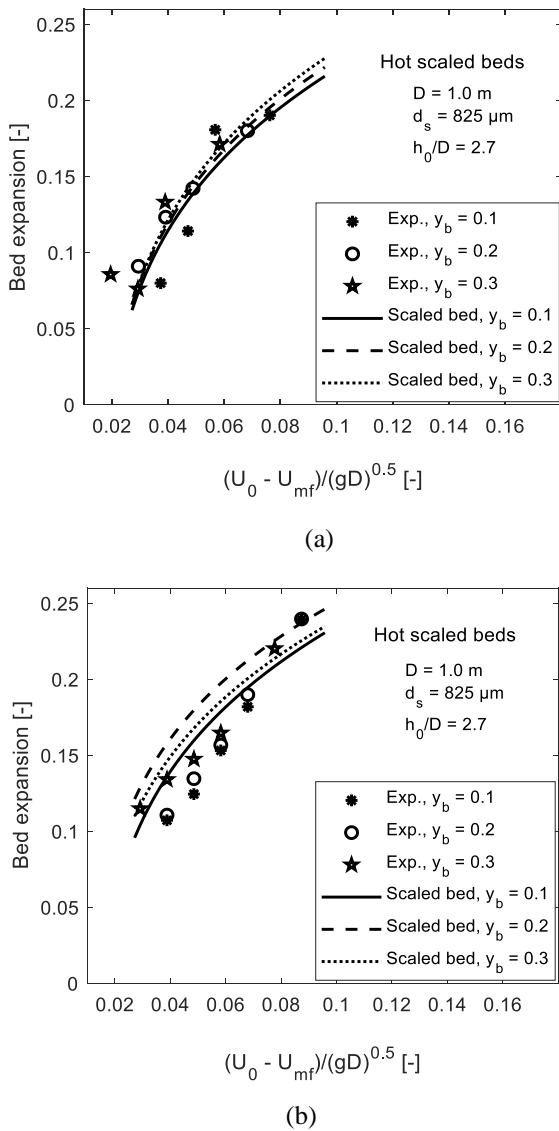
Figure 2. Bed expansion, comparing the bubbling behavior in the experimental small cold bed (diameter = 10.4 cm, and 293 µm sand particles) with that in the hot scaled bed of diameter 100 cm containing 827 µm sand particles at different biomass loads of (a) wood chips and (b) wood pellets.

The grouping of particles in the CPFD code makes the simulation faster, thereby increasing its application to industrial systems. Detailed descriptions of the CPFD model and its numerical scheme can be found in Chen et al (2013). The Barracuda software has been successfully used in different studies including investigation of fluid-particle behavior in a binary mixture of bed material and biomass particles (Fotovat et al, 2015b) and classification of particles of different sizes and densities (Chladek et al, 2018).

In implementing the CPFD model, the particle properties were the same for both biomass and sand particles as those used in the cold bed experiment (Agu et al, 2019a). Both beds ($D = 10.4$ cm and $D = 100$ cm) contain 20 vol% wood chips for each air velocity.

Table 1. Comparison of the dimensionless numbers between the small cold and the large hot beds.

Parameter	Unit	Small bed with 20 vol.% biomass	Large Bed with 20 vol.% biomass
T	°C	25	800
p	atm	1	1
D	m	0.104	1.0
ρ_g	kg/m^3	1.18	0.33
μ_g	Pa.s	1.8x10^{-5}	4.3x10^{-5}
d_s	µm	293	827
ρ_s	kg/m^3	2650	2650
U_{mf}	m/s	0.079	0.243
$U_{mf}^2/(gD)$	-	6x10^{-3}	6x10^{-3}
$\rho_g U_{mf} d_s /\mu_g$	-	1.54	1.53
Ar	-	2500	2600
h_0/D	-	4.8	2.7
h_0/d_s	-	1800	3300

The airflow rate at the bottom of the column was assumed uniform across the bed. There are different drag models in the Barracuda library. As the Wen and Yu (1966b) drag model has been used in many applications with good results, this model was also used for all simulations in this study. The default model settings, as given in the Barracuda 17.05 version, were applied.

The geometries are divided into uniform grids. The grid size for the small bed is 13 x 13 x 12.2 mm while that of the larger bed is 9.6 times larger. Relatively coarse grid sizes were used to accommodate the biomass particles (6.87 mm diameter) within each computational cell since in Barracuda it is difficult to capture particles larger than the grid size. Each bed was initialized as a uniform mixture containing 20% wood and 80% sand particles over the bed height. The biomass and sand particles were set as two different particle species to track the individual particle motion and location along the bed. The bottom part of the bed was set as the velocity flow boundary and the top as the pressure boundary. Four different superficial gas velocities were applied in each bed. The simulations were run for 20 s with an initial time step of 0.001 s. The fluid pressure and solid volume fraction at two different positions along the bed axis were used for further analysis.

Figure 3 (a and b) compares the distribution of the solid species (sand and biomass particles) between the two different beds from the initial well mixed state to the final state. For the same biomass volume fraction, the number of biomass particles per unit length is

proportional to the cross-sectional area of the bed. As a result, the scale on the colour bar reflects a higher biomass concentration in the hot bed compared to the cold bed, which is 100 times smaller in cross-sectional area. At the final state, the results show that particle distributions are similar in both beds; most of the biomass particles moved to the bed surface as similarly observed in the experiments. The pressure value at the two indicated positions in the different beds were used as the basis for Figure 3 (c), which shows the variation of pressure drop normalized with the mean value within the last 5 s of the simulations. As can be seen, the fluctuation of pressure drop is similar in both beds, indicating that bubbles rise at approximately the same frequency in the beds. The higher fluctuation amplitude in the hot bed is due to a larger pressure drop over the 70 cm between the measurement positions compared to 15 cm in the case of the cold bed. It should be noted that no scaling rule is applied in choosing the pressure data positions. However, the positions are such that one is within the grid zone (occupying up to 25% of the bed height) and the other within the constant density zone. The variation of mean pressure drop and pressure fluctuations at different gas velocities are shown in Figure 4.
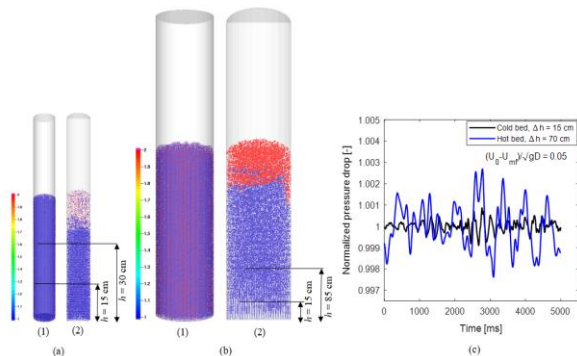
Figure 3. CPFD simulations showing solids distribution (a) cold bed of diameter 10.4 cm, height 50 cm with (1) initial state (2) final state (b) hot (800 ℃) scaled bed of diameter 100 cm, height 270 cm with (1) initial state (2) final state (c) pressure drop over the indicated positions within the last-5 s of the simulation time. On the colour scale, "1" indicates sand (blue), "2" indicates biomass (red) and a number in between indicates the particle mixture.

For the same Froude number, $Fr = (U_0 - U_{mf})/\sqrt{gD}$, Figure 4(a) shows that the pressure drop per unit length is close in the two beds. The approximately constant value of the pressure drop at different gas velocities indicates that the bed is well fluidized. Figure 4(b) gives the relative pressure fluctuation computed as σ/p, where σ is the standard deviation and p is the mean fluid pressure over the last 5 s of the simulation time. The value of σ/p increases with increasing gas velocity due to increasing bubble flux. Up to $Fr = 0.05$, the pressure fluctuations are similar in the two beds, and

also the same at both bed positions. At higher Froude numbers, the difference in the values of σ/p between the two beds (particularly at the upper position) suggests that slugs flow in the beds (although not clear in the simulations). In the cold bed, the experiments show that slugs begin to flow in the upper bed position (observed at 28.7 cm above the distributor) when $Fr = 0.078$. The flow of slugs results in higher fluctuations and bed instability.

With the strong agreement in both the simulated and experimental results, it therefore shows that by using the computed scaled bed aspect ratio, the bubble flow behavior in the hot large bed is similar to that in the cold small bed within the bubble regime.

(a)

(b)
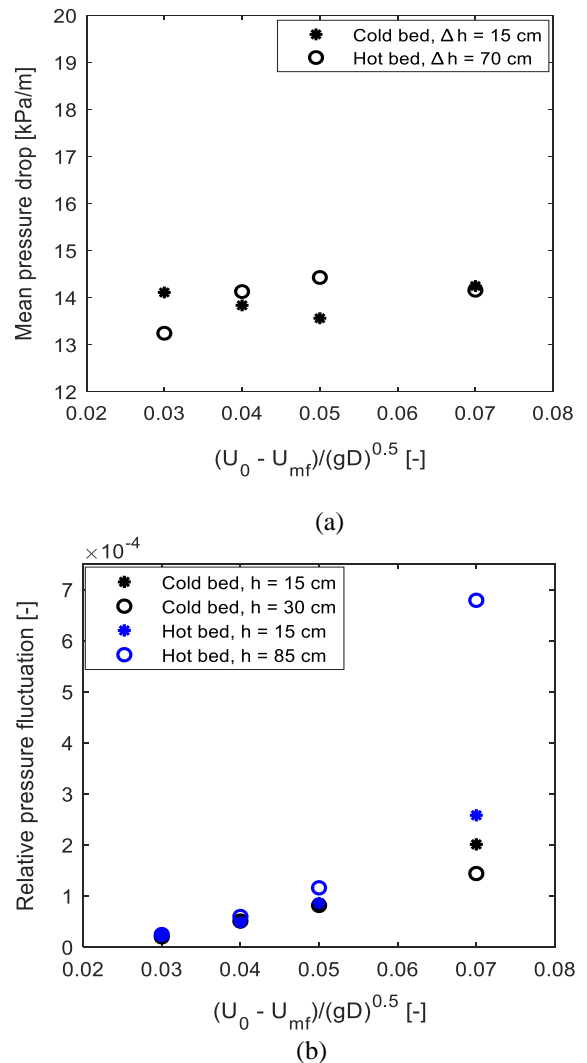
Figure 4. Comparison between cold bed of diameter 10.4 cm, aspect ratio 5 and hot (800 ℃) scaled bed of diameter 100 cm, aspect ratio 2.7 at different values of $(U_0 - U_{mf})/\sqrt{gD}$ (a) mean pressure drop (b) relative fluid pressure fluctuation.

4 Conclusions

In this study, a method for scaling up a cold bed to a hot larger bed was introduced. Using an output such as bed expansion ratio, the proposed scaling approach provides a means of achieving dynamic similarity between two different beds without constraining the fluid and particle properties as well as the bed height, thereby allowing the same properties of fuel particles and bed material type to be used. For illustrations, a 10.4 cm diameter cold bed containing a mixture of biomass and sand particles was scaled up to a large hot bed of 100 cm diameter and the results from CPFD model simulations showed that both beds have clear similarity in bubbling behavior.

Beside bed expansion ratio, other bubble properties including the rise velocity and volumetric bubble flux can also be considered in the scaling approach, provided that such properties are well correlated with the bed geometry and particle properties.

Although the new scaling up approach introduced in this work is verified numerically using commercial simulation software, further verifications using experimental data are necessary.

References

C. E. Agu, L.-A. Tokheim, C. Pfeifer and B. M. E. Moldestad. Behaviour of Biomass Particles in a Bubbling Fluidized Bed: A Comparison between Wood Pellets and Wood Chips. *Chemical Engineering Journal,* 363: 84-98, 2019(a).

C. E. Agu, A. Ugwu, C. Pfeifer, M. S. Eikeland, L.-A. Tokheim and B. M. E. Moldestad. Investigation of Bubbling Behaviour in Deep Fluidized Beds at Different Gas Velocities using Electrical Capacitance Tomography. *Industrial & Engineering Chemistry Research,* 58(5): 2084-2098, 2019(b).

C.E. Agu, L.-A. Tokheim, M. S. Eikeland and B.M.E. Moldestad. Improved models for predicting bubble velocity, bubble frequency and bed expansion in a bubbling fluidized bed. *Chemical Engineering Research and Design,* 141: 361-371, 2019(c).

C.E. Agu, C. Pfeifer, M. Eikeland, L.-A. Tokheim and B.M.E. Moldestad. Models for predicting average bubble diameter and volumetric bubble flux in deep fluidized beds. *Industrial & Engineering Chemistry Research,* 57(7): 2658-2669, 2018.

M.J. Andrews and P.J. O'Rourke. The multiphase particle-in-cell (MP-PIC) method for dense particulate flows. *International Journal of Multiphase Flow,* 22(2): 379-402, 1996.

M. Biglari, H. Liu, A. Elkamel and A. Lohi. Application of scaling-law and CFD modeling to hydrodynamics of circulating biomass fluidized bed gasifier. *Energies,* 9(7): 504-518, 2016.

C. Chen, J. Werther, S. Heinrich, H.-Y. Qi and E.-U. Hartge. CPFD simulation of circulating fluidized bed risers. *Powder Technology,* 235: 238-247, 2013.

J. Chladek, C.K. Jayarathna, B.M.E. Moldestad and L.-A. Tokheim. Fluidized bed classification of particles of different size and density. *Chemical Engineering Science,* 177: 151-162, 2018.

J. H. Choi, J. E. Son and S. D. Kim. Bubble size and frequency in gas fluidized beds. *Journal of Chemical Engineering of Japan,* 21(2): 171-178, 1988.

R. C. Darton, R. D. La Naueza, J. F. Davidson and D. Harrison. Bubble growth due to coalescence in fluidized beds. *Transactions of the Institution of Chemical Engineers,* 55: 274, 1977.

F. Fotovat, R. Ansart, M. Hemati, O. Simonin and J. Chaouki. Sand-assisted fluidization of large cylindrical and spherical biomass particles: Experiments and simulation. *Chemical Engineering Science,* 126: 543-559, 2015(a).

F. Fotovat, A. Abbasi, R.J. Spiteri, H. de Lasa and J.A. Chaouki. CPFD model for a bubbly biomass-sand fluidized bed. *Powder Technology,* 275: 39-50, 2015(b).

L. R. Glicksman, M. Hyre and K. Woloshun. Simplified scaling relationships for fluidized beds. *Powder Technology,* 77(2): 177-199, 1993.

L. R. Glicksman. Scaling relationships for fluidized beds. *Chemical Engineering Science,* 43(6): 1419-1421, 1988.

L. R. Glicksman. Scaling relationships for fluidized beds. *Chemical Engineering Science,* 39(9): 1373-1379, 1984.

A. Hepbasli. Estimation of bed expansions in a freely-bubbling three-dimensional gas-fluidized bed. *International Journal of Energy Research,* 22: 1365, 1998.

M. Horio, A. Nonaka, I. Sawa and Y. Muchi. A new similarity rule for fluidized bed scale-up. *AIChE Journal,* 32(9): 1466-1482, 1986.

S. Kraft, M. Kuba and H. Hofbauer. The behavior of biomass and char particles in a dual fluidized bed gasification system. *Powder Technology* 338: 887-897, 2018.

S. Mori and C. Y. Wen. Estimation of bubble diameter in gaseous fluidized beds. *AIChE Journal,* 21(1): 109-115, 1975.

J. B. Romero and L. N. Johanson. Factors affecting fluidized bed quality. *Chemical Engineering Progress. Symposium Series* 58: 28, 1962.

P. Salatino and R. Solimene. Mixing and segregation in fluidized bed thermochemical conversion of biomass. *Powder Technology* 316: 29, 2017.

R. K. Thapa and B. M. Halvorsen. Scaling of bubbling fluidized bed reactors with Glicksman's viscous limit set and CFD simulation. *International Journal of Computational Methods and Experimental Measurements* 2(2): 135-144, 2014.

C. Y. Wen and Y. H. Yu. A Generalized Method for Predicting the Minimum Fluidization Velocity. *AIChE Journal,* 12(3): 610-612, 1966(a).

C. Y. Wen and Y. H. Yu. Mechanics of fluidization. *Chemical Engineering Progress Symposium Series,* 162: 100-111, 1966(b).

J. Werther. Effect of gas distributor on the hydrodynamics of gas fluidized beds. *German Chemical Engineering,* 1: 166, 1978.

J. Zhang and G. Xu. Scale-up of bubbling fluidized beds with continuous particle flow based on particle-residence-time distribution. *Particuology,* 19: 155-163, 2015.

Challenges of CAD conversion to 3D development environments with respect to kinematic dependencies

Philipp Braun   Martin Sliwinski   Johannes Hinckeldeyn   Jochen Kreutzfeldt

Institute for Technical Logistics, Hamburg University of Technology, Germany, {philipp.braun, martin.sliwinski, johannes.hinckeldeyn, jochen.kreutzfeldt}@tuhh.de

Abstract

Due to the increasing utilization of 3D development environments for industrial use cases by emerging topics like virtual reality, new challenges in the conversion of computer-aided-design (CAD) models to 3D models have become visible. Particularly, the conversion of the kinematic constraints turns out to be complex and requires extensive manual efforts. This paper discusses these challenges with a focus on the problems in the conversion of kinematic constraints. Further, a new approach is proposed and validated regarding the automated conversion of the kinematic constraints of an exemplary CAD model. Thereby, this work contributes to an accelerated and simplified development process of industrial applications within 3D development environments.

Keywords: computer aided design (CAD), 3D, kinematic dependencies, CAD conversion, constraints

1   Introduction

CAD systems are indispensable in industrial product development processes and represent the state of the art in the area of product design software. However, product development tasks often go far beyond pure product design and require additional software solutions for simulation, visualization or other tasks. While CAD systems are the preferred tools for product development, 3D development environments[1] (3DDE) are used for visualization in high-quality renderings, videos or virtual reality applications (Davila Delgado et al., 2020). Because of the limited compatibility between CAD systems and 3DDE, file conversions are usually required and lead to a loss of information (e.g. parameters, materials, kinematic constraints) (Raposo et al., 2006). Manual restoration of lost information can be time consuming and must be repeated with every adjustment of the design. Even existing CAD conversion tools are not able to fully convert kinematic constraints (Stelzer et al., 2014). Other scientific publications identified this problem as well but could

not present a general applicable solution (Whyte et al., 2000). For that reason, this paper will present existing challenges regarding the conversion of kinematic constraints followed by an approach for an automatic conversion concept.

This paper is structured as follows: In section 2, the state of the art highlights current challenges regarding the conversion of CAD models to 3DDE compatible files. Further, problems of current approaches to overcome the existing challenges in the file conversion will be discussed. In section 3, an approach for a concept which addresses the identified issues is presented. A validation and discussion of the presented approach is given in section 4. This work closes with a conclusion in section 5.

2   State of the art

The following section focuses on the state of art and identified challenges in the conversion workflows from CAD models to files supported by 3DDE. First, currently available file formats will be discussed. Second, challenges regarding the conversion of kinematic constraints are presented. Finally, current approaches to overcome the conversion issues will be presented.

2.1   File formats for 3D-data exchange

Challenges in the conversion of 3D models from CAD systems to 3DDE arise, because most software systems use their own native file formats. In addition, the most software systems offer only limited capabilities to export or import data to other software. In order to overcome these issues, neutral exchange formats have been developed, which can generally be read by any common CAD- or 3D-system. However, the converted content often does not go beyond pure geometry (Dimitrov and Valchkova, 2011). The possibility to store additional information like materials, parameters, kinematic constraints etc. can vary from format to format. Furthermore, only a limited number of CAD systems offer the possibility to export these data.

---

[1] Popular examples are 3DsMax, Cinema4D, Blender or Unity.

One major difference between CAD and 3DDE models is the representation of surfaces. CAD-Models are described by non-uniform-rational-B-spline (NURBS) surfaces which are an exact mathematical representation of the geometry. The most commonly used CAD-exchange formats are STEP (.stp) and IGES (.igs) which are both neutral formats. In contrast, 3DDE use a discretized, polygon-based grid to generate surfaces as current 3DDE cannot render NURBS surfaces in real-time (Kumar and Manocha 1996; Sikorska 2008). When moving from CAD to a 3D application, polygonization (tessellation) of the CAD model is generally required. As a result, the used file formats differ fundamentally. In general, the tessellation process only translates surfaces belonging to the 3D model. Construction geometry like points, axes or planes are not considered and consequently lost during conversion. The most important neutral 3D-exchange formats are Wavefront OBJ (.obj), COLLADA (.dae), STL (.stl) and VRML (.wrl). Further, the proprietary formats Filmbox (.fbx) and 3DS (.3ds) are provided by Autodesk but are supported by many different 3DDE. An overview of the mentioned file formats is presented in Table 1.

Table 1: Overview of the most important CAD- and 3D-exchange file formats.

File format	Suffix	Type	Context
STEP	.stp	neutral	CAD
IGES	.igs	neutral	CAD
Filmbox	.fbx	proprietary	3D
3DS	.3ds	proprietary	3D
Wavefront OBJ	.obj	neutral[2]	3D
COLLADA	.dae	neutral	3D
STL	.stl	neutral	3D
VRML	.wrl	neutral	3D

2.2 Kinematic constraints in CAD and 3DDE

Kinematics are commonly used in the product development to test and illustrate motions of future products within the design process. The correct definition of kinematic constraints is further required for dynamic simulation (Engelson, 2000) or functional tests in a virtual environment. The creation of such kinematics in CAD is usually done by assigning geometric relationships and thus deliberately restricting degrees of freedom (DOF). Different geometries and relations can be used here. The most frequently used

geometries are primitives like points, axes and planes, which can be defined as congruent, parallel, angular, or with a distance from each other. The totality of all geometrical constraints of a CAD model component results in its respective DOF of movement, which defines the kinematics of the assembly.

In contrast, 3DDE offers special features, commonly called kinematic joints, to simulate special kinematic behaviors. Due to the mesh-based surface representation in 3DDE, these joints are normally defined by assigning coordinates and vectors and manually restricting the respective degrees of freedom. Thus, the definition of kinematic constraints is strongly differing from the geometry-based constraint definition in CAD, which makes the conversion of constraints from CAD to 3DDE challenging.

Summarizing, the two main issues of file conversion from CAD to 3DDE result from the existence of a variety of different file formats and the loss of data during file conversion which exceed pure geometry. Especially kinematic constraints are lost, as none of the prevailing 3D-supported file formats can fully support the preservation of kinematic constraints from CAD models. There are different approaches to overcome these issues which are presented in the following subsection.

2.3 Existing approaches for the automated conversion of CAD models

To overcome the issues introduced in subsection 2.2, several approaches have been developed with a focus on the conversion of CAD files to 3DDE. These approaches can be divided into three categories based on the types of data which are transferred from the CAD software to 3DDE:

1. Manual conversion and recreation in 3DDE
2. Automatic conversion of geometries and user-defined meta-data
3. Automatic conversion of geometries and kinematic constraints

In the manual approach, 3D computer graphics software is used to convert CAD models to 3D supported file formats including the geometries to represent a certain model. After the conversion to a 3D-supported file format, the files can be imported into a 3DDE. Since all information regarding the kinematic dependencies are lost, a high manual effort is required to remodel these information to enable realistic kinematic simulations of interactions and motions (Wolfartsberger, 2019). This approach is currently the common workflow in the development of applications which include representations of CAD files in 3DDE.

---

[2] The ASCII variant of the Wavefront OBJ format is neutral while the binary variant is proprietary.

Within this workflow, the remodeling needs to be performed with every update of the CAD model and the effort reoccurs evenly. As a consequence, the second and third category focus on the development of automated solutions for the conversion of CAD files.

Approaches in category 2 focus on the conversion of CAD parts and attached meta-data from several subsystems like product data management systems. Compared to previously described approaches, these ones generate two datasets within the conversion process (Gebert et al., 2017). The first dataset includes the geometries of the CAD model in a common file format. The second dataset stores information which are not related to geometry. This includes information like the hierarchy structures of assemblies in CAD models and pre-defined animations of kinematic motion sequences. A further possibility in these approaches is the integration of process knowledge like production information from additional subsystems (Górski, 2017). Compared to the first category, the second category of approaches can support the automatic integration of pre-defined animations, though the integration of kinematic constraints is still not supported. Thereby, motion sequences can be visualized but a realistic interaction with a certain 3D model still needs to be modeled manually in a 3DDE.

The third category focuses on the conversion of kinematic constraints from CAD models to 3D supported file formats to enable full realistic simulations and animations of kinematic motion sequences in real-time based on user-defined input values. For this reason, the approach of Lorenz et al. (2016) also starts with the automatic conversion of CAD models and pre-defined animations. To enable interactive kinematic motion sequences, the functionalities of mechanisms in certain CAD models has to be reproduced in the 3DDE. Therefore, the calculation of kinematic constraints is performed in a CAD system which is directly connected to the 3DDE via the programming interface. Using this connection, all manipulations of positions or orientations of parts in the 3DDE are streamed to the CAD system. Based on this transmitted information, the positions and orientations of all connected parts within the CAD model are recalculated. These recalculated data are streamed back to the 3DDE which enables a full interactive representation of CAD models in 3DDE (Andaluz et al., 2016). However, by connecting a 3DDE to a CAD system the real-time streaming between these tools is indispensable. Using this approach, a stand-alone solution for the conversion of kinematic constraints in a 3DDE is difficult to implement. Further, a reliable streaming highly dependents on the stability and performance of the connection.

In summary, none of the mentioned categories of approaches supports the full implementation of all constraints from CAD models in 3DDE as a stand-alone application. To implement a stand-alone application in a 3DDE, an automated remodeling of all kinematic constraints and DOF of the parts in certain CAD models is required. This remodeling must be implemented exclusively with tools provided by the 3DDE.

# 3 Approach for an automated conversion of kinematic constraints

This section presents an approach of a conversion concept from CAD systems to 3DDE with respect to kinematic constraints. The aim is to show a general and conceptual way of translating kinematic constraints from CAD to 3DDE not to present a fully functional software tool. The first subsection discusses prerequisites like the used file format and the basic conversion principle. Subsequently, the solution for the assembly hierarchy conversion and the conversion of constraints is presented. The origin of this work is based on an industrial use case in which a developed product had to be tested in a virtual environment. The design was created in Autodesk Inventor, while the visualization in 3D takes place in the 3DDE Unity. The following implementation of the conceptual approach applies basically to the conversion from Inventor to Unity, but can be adapted to further CAD systems.

3.1 Prerequisites for the conversion concept

The approach presented in this work uses the programming interfaces provided by Autodesk Inventor and Unity. The programming language used for Autodesk Inventor is VBA while Unity uses C#. Since the most CAD systems and 3DDE offer programming interfaces, the proposed solution can be adapted to other software combinations.

As shown in chapter 2.2, 3DDE like Unity only support polygon-based file formats for the 3D-data-exchange. For this work, we chose the .obj format which was selected primarily because of its easy processing and manipulation capabilities. In contrast to other file formats like .fbx and .dae, the .obj format can be natively exported from Autodesk Inventor. The disadvantage is the limited capability to store additional information. However, this work is not intended to present a final solution for the constraint problem, but to show a general way to transform kinematic constraints from CAD to 3DDE.

Due to the loss of hierarchy information, a manipulation of the .obj file is necessary to restore the local coordinate systems of the part, which is described more detailed in the following section. In addition to the .obj file, a text-based extension file is generated, which is used for storing missing information. This text file is automatically created within the CAD conversion and contains necessary information for every part of the assembly to restore the original structure. An illustration of the basic conversion concept is presented in figure 1. Two main steps can be differentiated here: the
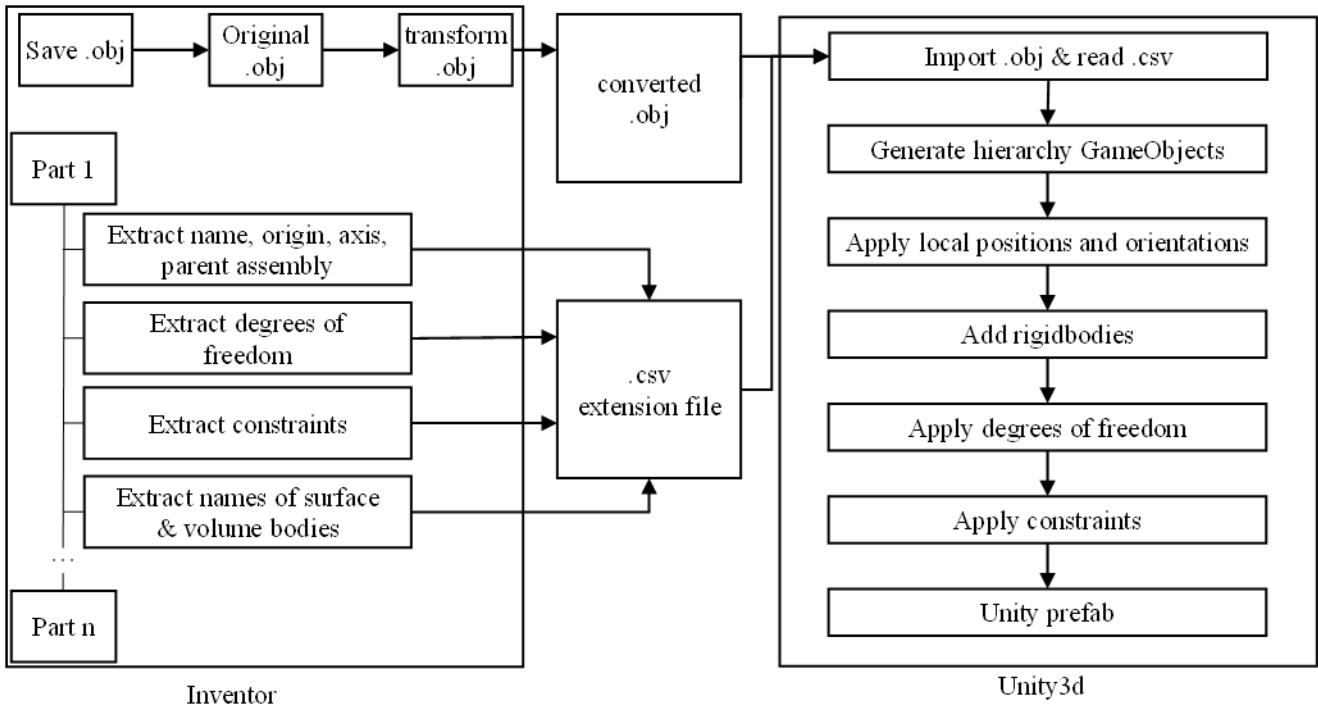
Figure 1: Software architecture of the proposed conversion approach.

conversion of the assembly hierarchy including all local coordinate systems and the conversion of the entire movement kinematics of the assembly.

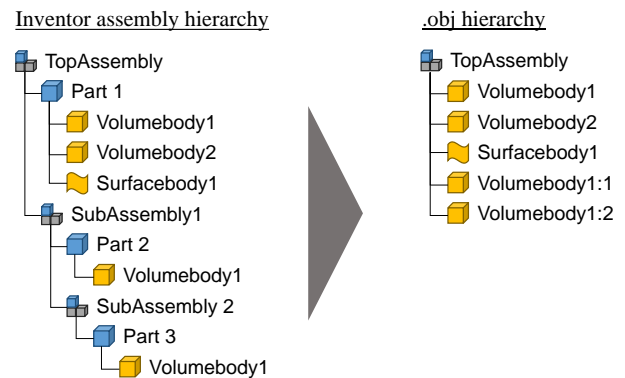## 3.2 Conversion of assembly hierarchy

A notable disadvantage of the .obj file format is the missing capability of preserving the assembly hierarchy structures. Thereby, the conversion leads to the consolidation of all volume and surface bodies on different hierarchy levels of an assembly resulting into one unstructured collection of bodies (see figure 2). Furthermore, the .obj format does not support the use of different coordinate systems, which means that all bodies are positioned with respect to a common origin point during the conversion.

Further, the volume and surface bodies[3] of the entire assembly are converted into a mesh-based geometry and written to the .obj file one-by-one during the conversion. If multi-body parts were used in the original design, this is no longer recognizable after the conversion and the individual bodies can no longer be assigned to the associated parts.

To correctly restore the hierarchy structure of the CAD model after the conversion, knowledge of the local coordinate systems of each part in relation to the global coordinate system is essential. The assignment of parts and subassemblies to the various hierarchical levels must also be known. For this reason, each part or sub-assembly of the main-assembly used in Autodesk

Inventor is investigated in a loop and the name, type, superior assembly as well as associated volume and surface bodies are written to the extension file. The name corresponds to the component part-name that was assigned in Inventor. The type indicates whether it is a part or an assembly and the higher-level assembly is required to restore the hierarchy levels.

In this approach, the lost parts and assemblies of the CAD model are represented by newly instantiated Unity GameObjects[4]. After instantiation, the volume and surface bodies from the CAD model are attached to the instantiated GameObjects representing their superordinate parts. After attaching volume and surface

Figure 2: Assembly hierarchy structure in CAD models and .obj files after conversion.

---

[3] A single CAD-part can have multiple volume and surface bodies. After the conversion to .obj and reimport the bodies appear as independent parts.

[4] GameObjects are the basic class of entities within a Unity virtual environment.

bodies to their respective GameObjects, the hierarchy structure of the CAD model has to be restored.

Due to the loss of the local coordinate systems during the conversion, the origin point of the bodies is no longer coincident with the original coordinate system origin. All vertices describing the body's mesh inside the .obj file are given relative to the global coordinate system. To correct this, a coordinate transformation of all vertices is required. This transformation takes place directly in the original .obj file where all bodies are moved to the global origin. To transform the coordinates, the individual transformation matrix of each part is needed. The values of this matrix can be evaluated in Inventor and are written to the extension file. After the transformation all vertices of a body are given in relation to the local coordinate system of the associated part. When importing this file to Unity, all local coordinate systems of the bodies are coincident with the global coordinate system. To relocate the individual bodies, the extension file contains the coordinates of the origin points and the direction vectors of the X-, Y- and Z-axis for each component in relation to the global coordinate system.

Using this information, the parts and assemblies can be relocated with correct coordinate systems in Unity. This is performed by a Unity extension script which reads the information regarding the local position and orientation of all parts from the extension file. In the next step, local positions and orientations are assigned to all GameObjects representing the parts and assemblies of the CAD model. Additionally, the assigned position and orientation values must be translated from the right-handed coordinate system of the CAD system to the left-handed coordinate system of Unity.

3.3 Conversion of kinematic dependencies

In the following subsection, the algorithm for converting kinematic dependencies, which is the core of this work, will be described. The basic concept of the algorithm is valid across different software tools and file formats. The implementation can be considered as a proof of concept.

In common CAD systems, the connection between multiple rigid bodies within an assembly design is realized by assigning kinematic constraints to pairs of rigid bodies. These constraints are based on primitive geometries like points, lines or planes, which can be defined as coincident, angular and distant (Haller et al., 2012). Most of the 3DDE offer only a limited functionality compared to CAD systems for assigning kinematic constraints. Instead, constraints are simulated using a physics engine in combination with special attributes assigned to the individual objects. In this work, two essential attributes of Unity (hinge joint and configurable joint) are used to remodel kinematic constraints. The hinge joint connects two bodies to form

a joint, restricting all translational and rotational DOF excluding the rotation around one joint axis. The joint axis is defined by an origin point and a direction vector. The configurable joint is a flexible component, which enables the creation of kinematic joints, but also the restriction and unlocking of certain DOF. This is performed by assigning an individual coordinate system to the specific configurable joint. Subsequently, all six DOF can be restricted or unlocked individually.

The algorithm for automatically converting the kinematic dependencies consists of four steps. The first and second steps must be performed in a CAD system and the last steps in a 3DDE:

1. Evaluation of DOF (CAD system)
2. Identification of rotational kinematic constraints (CAD system)
3. Reinstantiation of DOF (3DDE)
4. Setup of rotational kinematic constraints (3DDE)

In the first step, the DOF of all parts and assemblies within the CAD model are evaluated. Thus, the number of translational and rotational DOF as well as the respective coordinate directions needs to be determined. Any unrestricted DOF in combination with the corresponding direction vector is then written to the .csv file.

Subsequently, rotational constraints between different parts or assemblies within the CAD model are identified in step 2. As already described in the previous section, the connection between two parts is established by geometric relationships between two parts or assemblies within the CAD model. At the present early development stage, it was not possible to cover all possible combinations. Only constraints based on the congruence of two axes as a rotational axis are considered. Corresponding relationships are written into the extension file, stating the connected part or assembly and the axis direction.

The third step uses the input of the first step within the extension script for Unity to remodel the DOF of all parts and assemblies within the CAD model. Therefore, configurable joints are attached, corresponding to the information from the extension file, to restrict or unlock the DOF of the GameObjects representing the parts and assemblies of the CAD model.

Finally, components of the type hinge joint are applied to remodel the rotational constraints between different parts or assemblies of the CAD model within Unity. Therefore, the origin of the rotational axis is assigned to the hinge joint in the local coordinate system of the corresponding GameObject. Afterwards, the rotational axis is defined by the direction vector.

4 Validation and Discussion

The validation of the proposed approach for the automated conversion of kinematic dependencies from CAD models in Autodesk Inventor to the 3DDE Unity is performed using an exemplary CAD model of a mechanical assembly (Figure 3). Within the assembly, a rotational motion from a hand lever (A) is translated via a shaft to an eccentric panel (B). The bar (C) translates rotational motion to the slider (D) which has a single and linear direction of motion. The conversion of the CAD model was performed by the internal .obj converter of Autodesk Inventor and the developed extension script. To implement the DOF each GameObject representing a part or an assembly has two attached configurable joints. One is accountable for defining translational DOF and one for the rotational DOF. The rotation axis of A is modeled by a hinge joint as well as the centers of rotation on the top and bottom of C. The linear motion of D is modeled by unlocking the corresponding DOF within the configurable joint of the GameObject. The developed extension scripts are attached as a GitLab repository of the Hamburg University of Technology (Braun et al., 2020).

The presented simple and exemplary assembly was converted using the proposed approach including the rotational kinematic dependencies. This initially provides a proof of concept for the conversion concept with respect to the presented assembly. Currently, the proposed concept only considers constraints based on the congruence between two axes because these are primarily used for the generation of rotational DOF. Constraints based on further geometric relationships cannot be considered, yet. Translational DOF are not affected, since these are evaluated using the DOF analysis integrated in Invertor and do not require an analysis of the existing constraints.
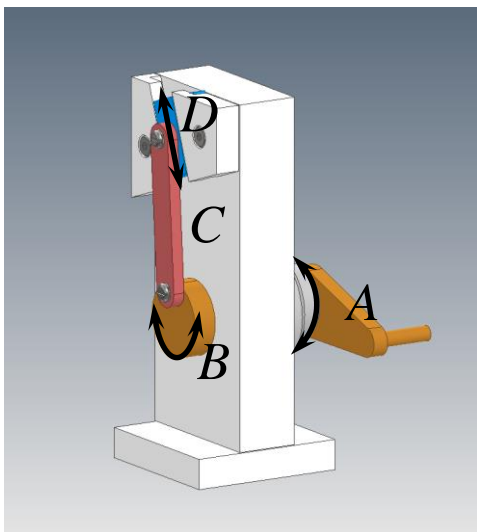
Figure 3: Exemplary CAD assembly for the validation of the proposed approach.

Future research activities should focus on algorithmic evaluation of the remaining combinations of constraints. Therefore, a more complex analysis of DOF is required (Lee-St.John, 2012). Especially, a possibility to find parts that are connected to each other as well as the type of their connection (translational / rotational) based on the existing DOF and the existing constraints of an assembly is of special interest. Additionally, subsequent research activities should focus on the used file formats. A detailed analysis is necessary to verify whether file formats like .fbx or .dae offer extended possibilities to transfer kinematic constraints of CAD models.

Despite the existing limitations, the presented approach can support the conversion of kinematic constraints in the application development process within 3DDE. This can lead to a faster and easier development of 3D applications for design reviews. Besides the easy visualization of kinematic dependencies, the automatic conversion supports the representation of results from simulations (Wolfartsberger et al., 2017).

5 Conclusion

This paper summarizes challenges in the conversion of CAD models to files supported by 3DDE. The challenges originate in the concurrent usage of CAD systems for mechanical designs and 3D development engines for high quality visualizations and application developments. Current file formats only support the conversion of geometries, but the conversion of kinematic constraints from CAD models to 3D files is not feasible. Therefore, different approaches from time-consuming manual restoration of kinematic constraints in 3DDE to the automated synchronization of CAD systems with 3DDE has been developed.

The conceptual approach for the conversion of kinematic dependencies of CAD models contributes to the standalone representation of CAD models in 3DDE. The current implementation enables the automated conversion of selected rotational kinematic constraints from Autodesk Inventor to Unity. Within the elaborated workflow, a file is generated to represent the geometries of the CAD model. An additional extension file contains missing information regarding the hierarchy structure, local coordinate systems, DOF and kinematic constraints. Using this information, the initial kinematic dependencies from the CAD model can be restored in Unity. In future research activities, it is planned to apply the proposed approach for a wider range of kinematic constraints.

References

V. H. Andaluz, F. A. Chicaiza, C. Gallardo, W. X. Quevedo, J. Varela, J. S. Sánchez, and O. Arteaga. Unity3D-MatLab Simulator in Real Time for Robotics Applications. In Lucio Tommaso de Paolis and Antonio Mongelli, editors, *Augmented Reality, Virtual Reality, and Computer*

Graphics, pages 246–263. Springer International Publishing, 2016. ISBN 978-3-319-40620-6. doi: 10.1007/978-3-319-40621-3_19.

P. Braun, M. Sliwinski, J. Hinckeldeyn, and J. Kreutzfeldt. *CAD to 3D Converter*. Hamburg University of Technology. Hamburg, 2020. Available online at https://collaborating.tuhh.de/w-6/forschung/depot-projekt/cad-to-3d-converter, checked on 07 July, 2020.

J. M. Davila Delgado, L. Oyedele, P. Demian, and T. Beach. A research agenda for augmented and virtual reality in architecture, engineering and construction. *Advanced Engineering Informatics*, 45 (3):101–122, 2020. doi: 10.1016/j.aei.2020.101122.

L. Dimitrov and F. Valchkova. Problems with 3D data exchange between CAD systems using neutal formats. *Proceedings in Manufacturing Systems*, 6 (3):127–130, 2011.

V. Engelson. *Kinematic information formats. Standards applicable for mechanical model translation from CAD to Modelica*. MathCore AB. Linköping, Sweden, 2000.

M. Gebert, W. Steger, R. Stelzer, and K. Bertelmann. Meta-model for VR-based design reviews. In *Proceedings of the 21st International Conference on Engineering Design (ICED17), Vol. 4: Design Methods and Tools, International Conference on Engineering Design, 21-25 August, 2017, Vancouver, Canada*, pages 337–346, 2017.

F. Górski. Building Virtual Reality Applications for Engineering with Knowledge-Based Approach. *Management and Production Engineering Review*, 8 (4):64–73, 2017. doi: 10.1515/mper-2017-0037.

K. Haller, A. Lee-St.John, M. Sitharam, I. Streinu, and N. White. Body-and-cad geometric constraint systems. *Computational Geometry*, 45 (8):385–405, 2012. doi: 10.1016/j.comgeo.2010.06.003.

A. Lee-St.John. Kinematic Joint Recognition in CAD Constraint Systems. In *Proceedings of 24th Canadian Conference on Computational Geometry, Canadian Conference on Computational Geometry, 8-10 August, 2012, Charlottetown, Canada*, pages 173–178, 2012.

M. Lorenz, M. Spranger, T. Riedel, F. Pürzel, V. Wittstock, and P. Klimant. CAD to VR – A Methodology for the Automated Conversion of Kinematic CAD Models to Virtual Reality. *Procedia CIRP*, 41 (3):358–363, 2016. doi: 10.1016/j.procir.2015.12.115.

A. Raposo, E. T. L. Corseuil, G. N. Wagner, I. H. F. dos Santos, and M. Gattass. Towards the use of cad models in VR applications. In *Proceedings of the 2006 ACM international conference on Virtual reality continuum and its applications, VRCIA06: ACM International Conference on Virtual Reality Continuum and Its Applications 2006, 14-17 June, 2006, Hong Kong, China*, pages 67–74, 2006. doi: 10.1145/1128923.1128935.

R. Stelzer, E. Steindecker, S. Arndt, and W. Steger. Expanding VRPN to Tasks in Virtual Engineering. In *Volume 1B: 34th Computers and Information in Engineering Conference, ASME 2014 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, 17-20 August, 2014, Buffalo, USA*, pages 1–8, 2014. doi: 10.1115/DETC2014-34277.

J. Whyte, N. Bouchlaghem, A. Thorpe, and R. McCaffer. From CAD to virtual reality: modelling approaches, data exchange and interactive 3D building design tools. *Automation in Construction*, 10 (1):43–55, 2000. doi: 10.1016/S0926-5805(99)00012-6.

J. Wolfartsberger. Analyzing the potential of Virtual Reality for engineering design review. *Automation in Construction*, 104:27–37, 2019. doi: 10.1016/j.autcon.2019.03.018.

J. Wolfartsberger, J. Zenisek, C. Sievi, and M. Silmbroth. A virtual reality supported 3D environment for engineering design review. In *Proceedings of the 2017 23rd International Conference on Virtual Systems and Multimedia (VSMM), International Conference on Virtual Systems and Multimedia, 31 October - 4 November 2017, Dublin, Ireland*, pages 1–8, 2017. doi: 10.1109/VSMM.2017.8346288.

ASME B16.9 Piping Components:
An Analysis of the Critical Dimensions

Gwen Anderson[1] Ida Helene Gjerde[1] Sondre Luca Helgesen[2] Mladen Jemenica[1] Rajan Kumar Thapa[1]

[1] Department of Process, Energy and Environmental Technology, University of South-Eastern Norway, Norway
`gwendolyn.anderson@outlook.com; idahbgjerde@gmail.com; {mladen.jecmenica,`
`rajan.k.thapa}@usn.no`
[2]Stressman Engineering AS, Norway `slh@stressman.no`

Abstract

Pipe components such as elbow, tees, and weldolets are produced at various places in the world. Although the component is produced in accordance with standard ASME (American Society of Mechanical Engineers) B16.9, measurements show variation in geometry. In piping component design, the geometries are assumed as perfect in accordance with the producer's specifications. This is a conservative approach in design that results in incorrect dimensioning of the components. Moreover, components made of duplex materials can have failure due to Hydrogen Induced Cracking (HISC) especially in subsea installations.

In this article, a series of measurements were taken for ASME B16.9 long radius elbows, utilizing an ultrasonic thickness gauge. The results are compared with the nominal straight pipe dimensions in ASME. The deviation is taken as a basis for Finite Element analysis.

A model of the elbow is developed using SolidWorks and simulated using ANSYS. The results compare the strength of the pipe fittings against their straight pipe counterparts with regard to internal pressure. This gives an overview of the deviations and utilization based on the measurements taken against the standard ASME B16.9 geometry.

Keywords: ASMEB16.9, piping, pipe bends, pipe elbows, FEM, standard components, ultrasonic thickness gauge

1 Introduction

HISC is a form of cracking caused due to high concentration of hydrogen in materials which leads to high amounts of stress at a concentrated area and can cause the pipe component to burst. This occurs when hydrogen diffuses into the material structure and changes the material's microstructure. The cracking tends to form parallel to the surface or the same direction as hoop stress. The structure will not immediately give in, but over time when being exposed to both strain and tensile stress, the pipe will suffer from internal cracking (Inspectioneering, 2020).

Pipe components used in the subsea industry are frequently exposed to HISC which weakens the component. Guidelines for avoidance of HISC is given in DNVGL-RP-F112 (Veritas, 2018) whilst ASME 16.9 (Engineers, ASME B16.9, 2012) requires the structural integrity for standard butt-welded components. Are all pipe components according to standard, and is the standard comprehensive enough? This question gives a background for why reevaluating ASME B16.9 is relevant. The Standard requires the pipe components to be of equivalent strength to its straight pipe counterpart. It is questioned if it is possible to confirm this statement, say something about the lower bound values for analysis and map eventual critical dimensions.

Details of the standard pipe bend components are evaluated by measuring the wall thickness using an ultrasonic thickness gauge. The data collected is evaluated both statistically as well as modeled in SOLIDWORKS and simulated using the Finite Element Method (FEM).

2 Pipe measurements

This article was started by mapping of critical dimensions of pipe components such as bends. Data was collected during a four-day visit to Scandinavian Flanges and Fittings (SFF) in Sandnes, Norway. The goal is to define a limitation for wall thickness in the form of a specific percentage or value.

2.1 Measurement equipment

The equipment used to take measurements was the Ultrasonic thickness gauge model MG26 from Panametrics-NDT. This gauge has a thickness range of 0.05 mm to 508 mm and a resolution of 10 micrometers. Ultrasonic rays do not have a good connection through the air, so the liquid gel is required to make proper contact with the component surface. The soap, Zalo was used as the gel. At the time of measurement, materials were not temperature controlled. According to Yr.no, temperatures in Sandnes, were between 2-8° C. The density of the material's grain spacing effects how the ultrasonic rays penetrates the material. Therefore, some materials are more difficult to measure than others.

2.2 Measurement locations

To have consistency in the data collecting process, the pipe bend was divided into four sections along the

curvature. These were labeled from Plane 1-5 including the ends of the bends. Figure 1 shows the planes in the order that they were sectioned, with an interval of 22.5°. All measurements refer to these planes in this order and position along the pipe bend.
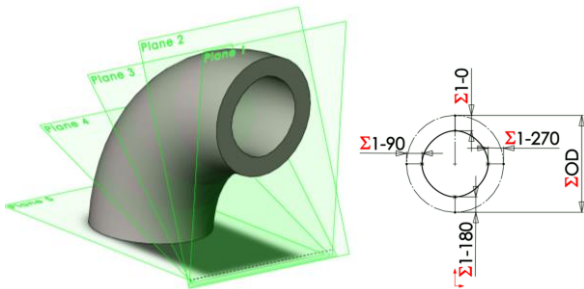
Figure 1. Measurement locations depicting the different planes along the pipe and the cross-section

At each plane, the cross-section was divided into many parts with fixed intervals to allow as many locations as needed. For the location of 0°, the extrados of the bend was chosen. This puts the intrados at 180°. Figure 2 shows the measurement locations and reading.

Figure 2. Measuring of 270 on plane 1 [left] and reading from NPS 6 SCH 80 Bend 1 [right]

2.3 Measurement process

Initially, extra measurements were taken on a single pipe bend, NPS 2 SCH 160, to create a large database of wall thickness. This included taking measurements at 45-degree increments for 8 locations per cross-section which gives a total 40 measurements for the bend. The number of measurements taken on each pipe was reduced after the NPS 2 was completed.

The data points were organized in Excel with each pipe size having its own sheet. The nominal wall thickness for an NPS 2 SCH 160 pipe is 8.378 mm. Accounting for the minus Mill Tolerance, the wall thickness is then 7.646 mm. In addition to wall thickness, using dial calipers the inner and outer diameter, and inner and outer bend radius were measured. Using the heat number on the stamp, details from the certified material test report were noted. A check for orality was performed on both pipe ends using dial calipers and noted if the measurement varied.

3 Modeling

The drawing programs SOLIDWORKS is used to make the geometry model of the component. The software program uses the centerline and cross-section profile to create a 3D model. The pipe model includes a perfectly circular outer diameter and a constant width for an ideal inner cross-section. To simulate the weakest allowed product that a manufacturer can produce, the NPS dimensions are used with the manufacturer's minus tolerance. This is specified in ASME B31.3 304.1.1.

3.1 Finite Element Analysis (FEA)

The SOLIDWORKS model was created using design table and global variables. The radius and outer diameter were set as global dimensions. The design table includes all the attributes and dimensions that are not defined as global from the configurations in Excel. The wall thickness was adjusted using the design table, which allowed for multiple model configurations.

Four split lines were added to the inner surface of the models at plane 2, plane 3, plane 4, and in the middle of the straight pipe as shown in Figure 3. The split lines allow for the simulation to extract results at the same location to ensure consistency in the evaluation.
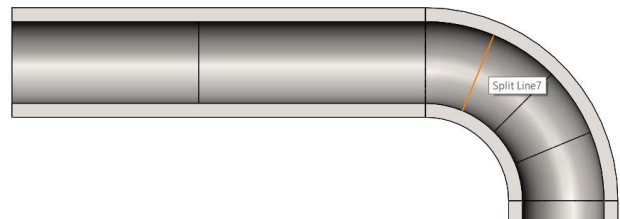
Figure 3. Split lines in the pipe model Simulation

Typically, a pipe is analyzed as a beam element (1D) with constant wall thickness. This creates a primary axis for movement and the other two are secondary directions. This model evaluates the complexity of the change in wall thickness of the pipe bend. The model was run as a solid (3D – tetrahedral element), which requires a larger mathematical model, and is not typically done for pipes.

3.1.1 Boundary Conditions

The straight pipes connecting the two ends of the bend were set as fixed at both ends. This avoids end cap effect as it restricts the movement on these surfaces completely. To reduce the concentrated stress from the fixed boundary on the attached straight pipe, various lengths were checked, and it was determined that the length of the pipe's bend radius was enough. In most cases 1.5 times bend radius was used for the straight pipe length. Figure 4 shows the fixed surfaces at the boundaries.
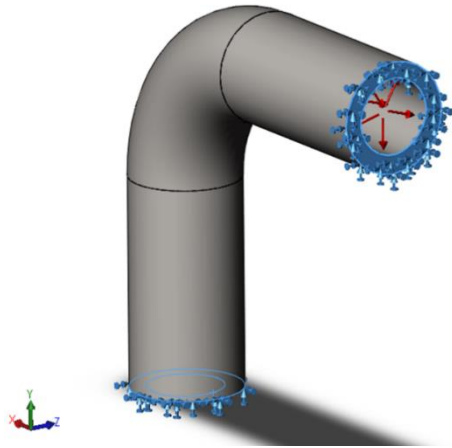
Figure 4. Pipe boundary conditions: fixed at both ends

3.1.2 Pressure

The internal pressure applied to all internal surfaces was determined using ASME B31.3 304.1.2 eq. (3a) (Engineers, ASME B31.3, 2002)

$$t = \frac{PD}{2(SE + \gamma P)} \tag{1}$$

By solving for pressure, P, the internal pressure is determined. This is calculated with an allowable yield strength of 440 MPa (S, UNS 32760 (SA-790) ASME BPVC II. D. M.-2015), outer diameter of 60.3 mm (D), wall thickness minus mill tolerance 7.6458 (t), coefficient for t>D/6 and ferric steel temperature lower than 482 °C of 0.4(γ, table 304.1.1), quality factor of 1 (E, table A-1B, spec no. A 790 material). For an NPS 2 SCH 160 pipe, the maximum pressure is 124.18 MPa.

3.1.3 Meshing

The pipe, NPS 2 SCH 160 was evaluated with a variety of mesh sizes. Using the split lines shown in Figure 3, results were compared for 4 different mesh sizes (15 mm, 8 mm, 2.5 mm, 1 mm). Figure 5 shows a comparison of the various mesh sizes. The medium(2.5 mm) mesh was used in further cases.
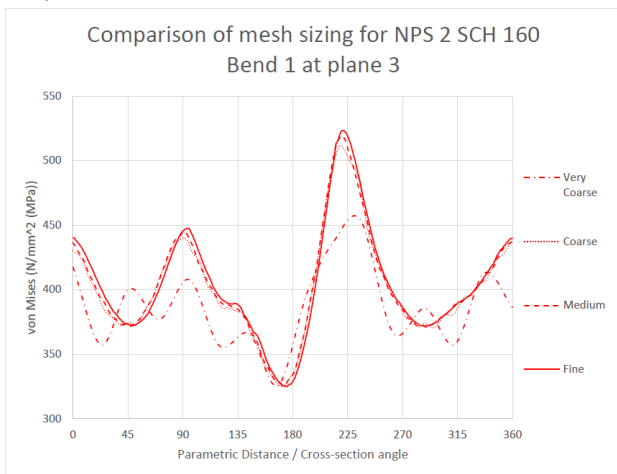

Figure 5. Probe results of plane 3 with variations on element sizes

4 Results and discussion

4.1 Measurement results

4.1.1 Results for NPS 2 SCH160 bends

For NPS 2 SCH 160, 10 bends were measured at 40 points. This gives a total of 400 measurements. Table 1 shows the measurement results for wall thickness at 90° increments at the cross-section and Table 2 shows the results of the measurements at the midpoint of a cross-section.

Table 1. Wall thickness at 90° increments

Bend NPS 2 SCH160	Wall thickness (mm)				
		Plane 1	Plane 2	Plane 3	Plane 4	Plane 5
	0	8.50	8.25	8.15	8.20	8.47
	90	9.25	9.10	8.97	9.15	9.21
	180	10.24	12.31	12.50	11.88	11.11
	270	9.67	9.43	9.18	9.47	9.57

Table 2. Wall thickness at midpoints of the 90° increments at the cross section

Bend NPS 2 SCH160	Wall thickness (mm) - Midpoints				
		Plane 1	Plane 2	Plane 3	Plane 4	Plane 5
	45	9.40	8.44	8.33	8.61	8.95
	135	9.97	10.28	10.20	9.70	9.63
	225	10.80	11.34	11.20	11.06	10.65
	315	9.16	8.94	8.80	8.80	9.12

In addition to wall thickness, using dial calipers the inner and outer diameter, and bend radius were measured which is shown in Table 3. Using the heat number on the stamp, details from the certified material test report were noted and presented in Table 4.

Table 3. Data collected for an NPS 2 SCH 160 bend: general dimension control

Dimension control (mm)			
	Plane 1	Plane 3	Plane 5
ID	42.40	N/A	42.00
OD	61.95	59.00	62.00
R_i	44.10	N/A	44.00
R_o	105.00	104.00	107.00

Table 4. NPS 2 SCH 160 bend: material test

CERTIFIED MATERIAL TEST REPORT			
Manufacturer	Shulz	Material	SA790
Cr	25.4	Standard	ASTM 790
Yield 0,2% (MPa)	576.0		
Yield 1% (MPa)	669.0	Origin	Spain
Tensile (MPa)	870.0	Manufactured	Germany

The measurements were analyzed using a sample standard deviation calculation. For a confidence of 95%, the Z-value used was 1.96. Assuming the pipe to be uniform in thickness, all measurements were weighted equally. This gives an average of 9.53 mm, 95% CI [9.44 mm, 9.61mm]. Figure 6 shows a normal distribution curve of NPS 2 SCH 160.
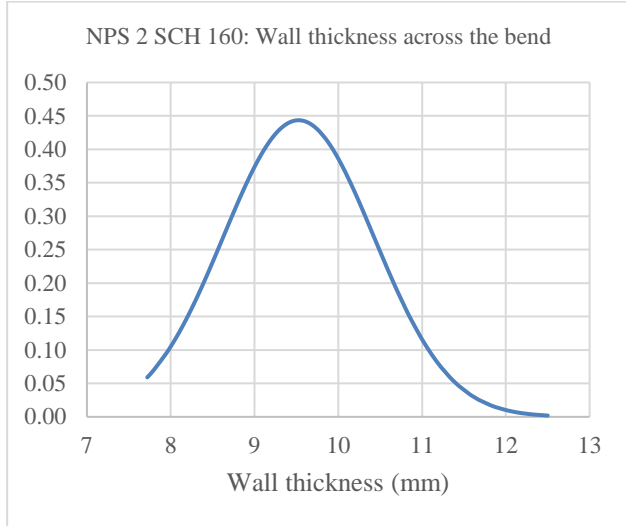

Figure 6. NPS 2 SCH 160, all measurements

There is a general trend of thinning along the extrados and thickening along the intrados. These three locations of wall thickness were the focus of the initial measuring process: end of pipe (plane 1 and plane 5), intrados (180/plane 3), and extrados (0/plane 3). These measurements were isolated and statistically evaluated to see the variation against the nominal wall thickness as shown in Figure 7.
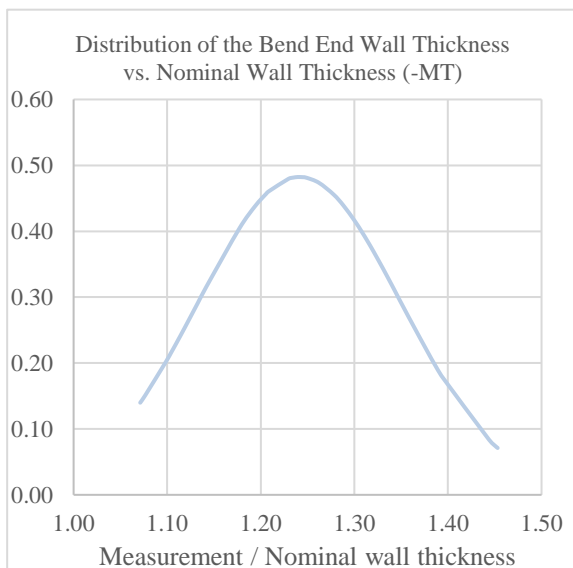

Figure 7. NPS 2 SCH 160 standard deviation diagram for the end wall thickness

4.1.2 Measurement results for the other pipe sizes

In total, 8 sizes of pipe bends were measured. Table 5 and Table 6 give an overview of all measurements.

Table 5. Quantity of pipes measured at SFF[1]

	NPS 2	NPS 3	NPS 4	NPS 6	NPS 8
SCH 80	7		7	10	
SCH 120					6
SCH 160	10	19	14	8	

Table 6. Confidence interval of the pipe measurements[2]

| Size	Intrados	Extrados	End
NPS 2 SCH 160	11.86 mm, 95% CI [11.60 mm, 12.12 mm]	8.42 mm, 95% CI [8.15 mm, 8.70 mm]	9.47 mm, 95% CI [9.38 mm, 9.56 mm]
NPS 2 SCH 80	8.08 mm, 95% CI [7.74 mm, 8.41 mm]	5.40 mm, 95% CI [5.19 mm, 5.61 mm]	6.76 mm, 95% CI [6.48 mm, 7.05 mm]
NPS 3 SCH 160	15.60 mm, 95% CI [15.23 mm, 15.98 mm]	11.79 mm, 95% CI [11.37 mm, 12.21 mm]	13.59 mm, 95% CI [13.4 mm, 13.79 mm]
NPS 4 SCH 160	17.79 mm, 95% CI [17.55 mm, 18.03 mm]	12.90 mm, 95% CI [12.69 mm, 13.11 mm]	15.14 mm, 95% CI [14.99 mm, 15.29 mm]
NPS 4 SCH 80	10.43 mm, 95% CI [10.15 mm, 10.71 mm]	7.95 mm, 95% CI [7.83 mm, 8.08 mm]	9.17 mm, 95% CI [9.03 mm, 9.30 mm]
NPS 6 SCH 160	23.47 mm, 95% CI [23.21 mm, 23.72 mm]	18.44 mm, 95% CI [18.17 mm, 18.70 mm]	20.72 mm, 95% CI [20.56 mm, 20.88 mm]
NPS 6 SCH 80	18.46 mm, 95% CI [18.19 mm, 18.72 mm]	14.53 mm, 95% CI [14.41 mm, 14.65 mm]	16.50 mm, 95% CI [16.32 mm, 16.68 mm]
NPS 8 SCH 120	22.66 mm, 95% CI [22.44 mm, 22.89 mm]	17.16 mm, 95% CI [16.96 mm, 17.36 mm]	20.10 mm, 95% CI [19.9 mm, 20.29 mm]

To compare the variation, all thicknesses were scaled against their respective nominal wall thickness. This creates a baseline of 1 to compare all pipes simultaneously in a single graph, where 1 represents the measured wall thickness equal to the nominal wall thickness minus MT. Figure 8(intrados), Figure 9(extrados), and Figure 10(end) all have the same trends of clumping of the normal distribution curves with an outlier being the NPS 6 SCH 80. There is a large spread in some of the sizes, NPS 3 SCH 160, for example. This had 19 total bends, so it had the largest sample size. Three different manufacturers were included in this sampling and 4 different heat numbers. Most of the bends measured had only one manufacturer and one heat number.
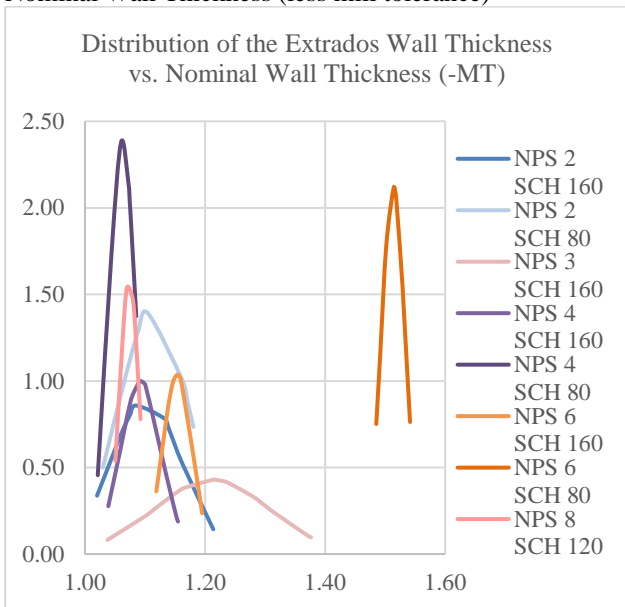
[1] NPS 8 SCH 120 is a seam welded pipe

[2] The bend end has four points for each bend as shown in Table 6. These are at plane 1 and plane 5 at 0° and 180°.
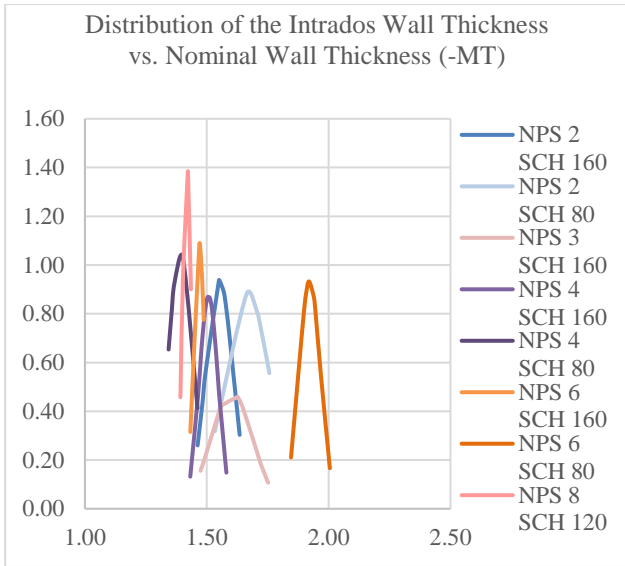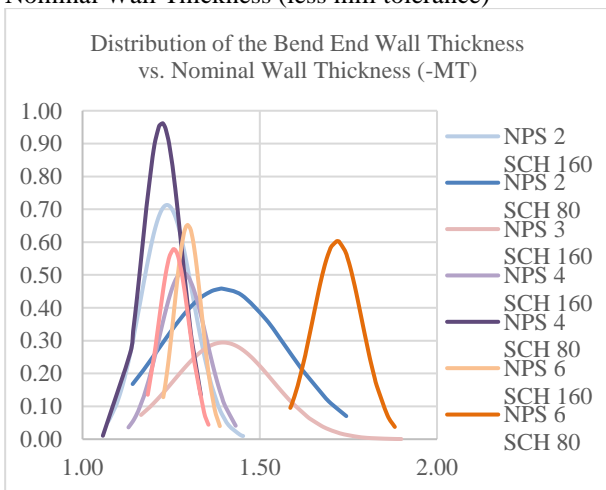
Figure 8. Distribution of the Intrados Wall Thickness / Nominal Wall Thickness (less mill tolerance)

Figure 9. Distribution of the Extrados Wall Thickness / Nominal Wall Thickness (less mill tolerance)

Figure 10. Distribution of the End Wall Thickness / Nominal Wall Thickness (less mill tolerance)

While mill tolerances are removed for comparison for a stress analysis, the standard pipe bends may be out of allowable tolerances due to total weight. Further research/observations are necessary as well as verification of the relevant standard for Super Duplex.

This article does not verify if the implication that the bends that were measured could be over the allowable limits due to the increased wall thickness. This can affect the total weight of the system and therefore support calculations and stress analysis. The NPS 6 pipe had nearly 50% more material. This is an outlier for the data set, but it should be noted that the pipe measurements included various sizes from the same manufacturer as the outlier size.

4.2 Simulation results

4.2.1 Comparison of quantity of measurements per bend

Two measurements were analyzed to evaluate critical aspects of the pipe bend when modeled and pressure is applied. This article compares the NPS 2 SCH 160 bend with 40 measurements and 20 measurements. The difference being the number of measurements per cross-section is shown in Figure 11.
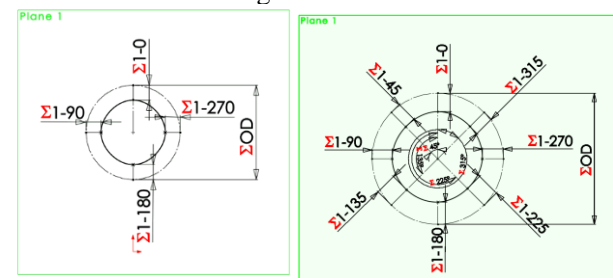
Figure 11. Cross section of plane on model 1. 90° increments for 20 in total [left] and 45° increments for 40 in total [right]
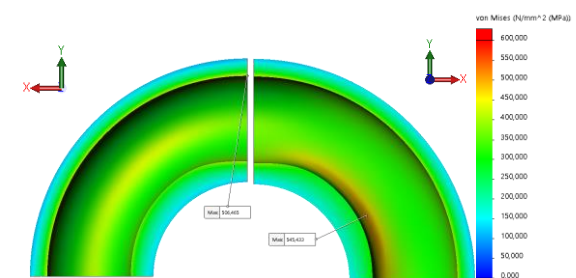
Figure 12. Stress of the pipe with 20 measurements [left] and 40 measurements [right]

Figure 12. shows the stress plot of NPS 2 SCH 160 bend. A transition of geometry, for example from a bend to a straight line, can have a stress concentration as seen in the result at the top of the bend on the pipe with 20 measurements, left in Figure 12. The split lines are utilized for a numerical comparison between the two results.

4.2.2 Comparison of measured models to standard pipe bends

The split line added in the middle of the attached straight pipe is used to check the strength of the standard straight pipe. The allowable minus mill tolerances are used in this analysis because all measured pipe bends were finished fabricated and the maximum stress the thinnest allowed straight pipe wall could tolerate is used as a baseline.

Figure 13. Straight pipe with constant wall thickness stress plot

Figure 13 shows the pipe stress from the applied internal pressure. The peak stress on the split line is 490 MPa. This was plotted as a baseline (black) in Figure 14. Note that the peak stress found from the model of 20 data points, at 484 MPa is lower than the straight pipe results, 490 MPa.

Figure 14. Varying quantity of measurements along a bend

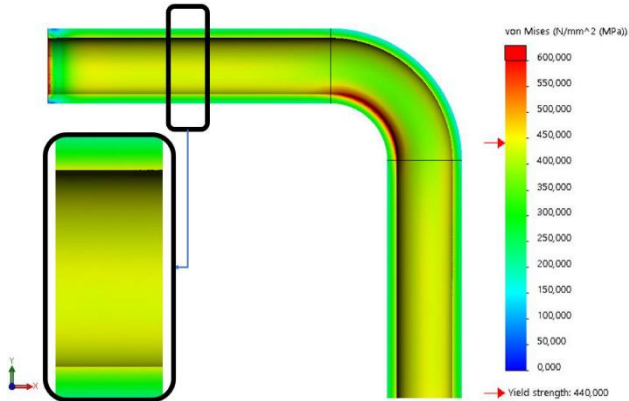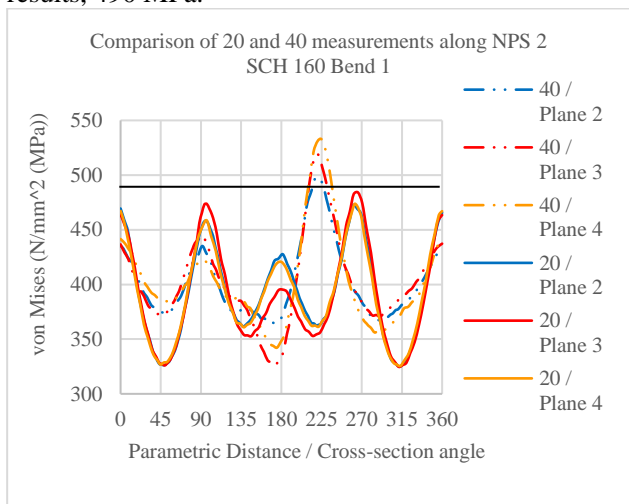The highest stress from the model with 20 measurements has peaks around 90° and 270° in Figure 14. The peak stress is 484 MPa on plane 3 at about 270 °. This correlates to the yellow area in Figure 12. The highest stress from the model with 40 measurements has peaks around 225° in Figure 14. The peak stress from the three split lines is on plane 4 at about 225° with 532 MPa. The peak stress on

plane 3 is 519MPa and plane 2 is 498 MPa, also at about 225 °. This correlates to the orange area in Figure 12.

The results of the highest stress within the bend (intrados) is compared with its respective straight pipe with the same pressure applied. To allow for comparison of all measured pipes, the results are in the form of a percent of the stress and wall thickness of a nominal straight pipe. The chart in Figure 15 shows the different measured pipe sizes. The vertical axis gives a comparison of the pipe bends stress and its straight pipe counterpart. The nominal pipe sizes that stay below the 100% line indicates that the bend has less stress than the straight pipe. The ones that are placed above the line contains more stress and are therefore weaker than the straight pipe. The horizontal axis marks the percentage of the measured wall thickness divided by the nominal pipe size i.e. the NPS 3 SCH 160 that is 65-73% thicker.

The NPS 4 SCH 80 has a higher stress than the straight pipe. This can be an issue because a pipe component, bend in this case, is supposed to be stronger than its straight pipe counterpart. Figure 15 shows that the thickness for this pipe size has its curve farthest left, implying thinner wall thickness than other sizes.
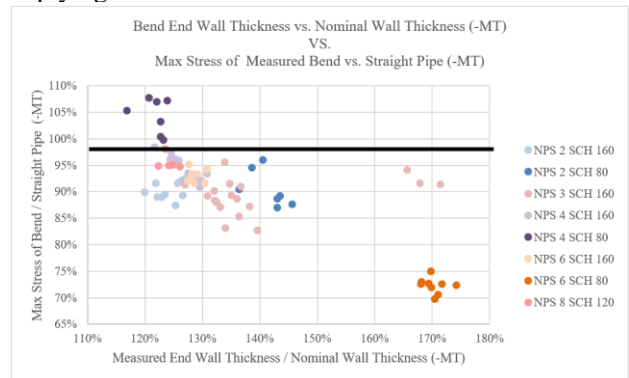
Figure 15. FEA Stress Results of the Bend End: A Comparison between the measured thickness and nominal thickness (-MT)
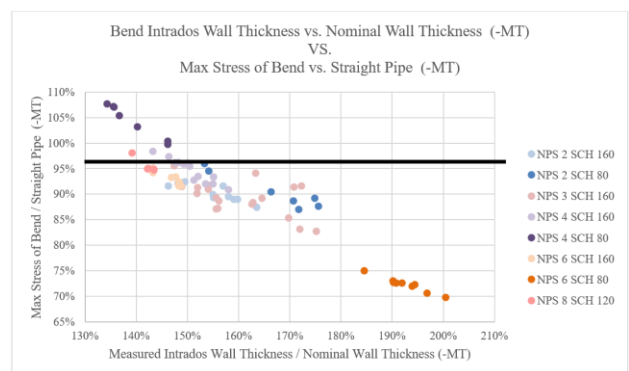
Figure 16. FEA Stress Results of the Intrados: A Comparison between the measured thickness and nominal thickness (-MT)
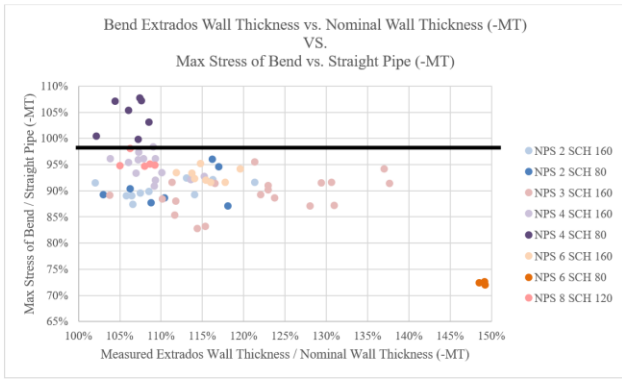
Figure 17. FEA Stress Results of the Extrados: A Comparison between the measured thickness and nominal thickness (-MT)

As shown in Figure 15, Figure 16, and Figure 17, there is a random spread in the different sizes. In Figure 16, an even curve is clearly formed which indicates a relation between the thickness of the intrados and the nominal pipe size when pressure is applied. The warehouse at SFF did not have any duplex bends in NPS 3 SCH 80, so 316 was also measured to include the size. Figure 18. shows the results of all measured bends, so there is a blend of materials. The linear trend line for the results of the intrados has an R^2 value of 0.8281.
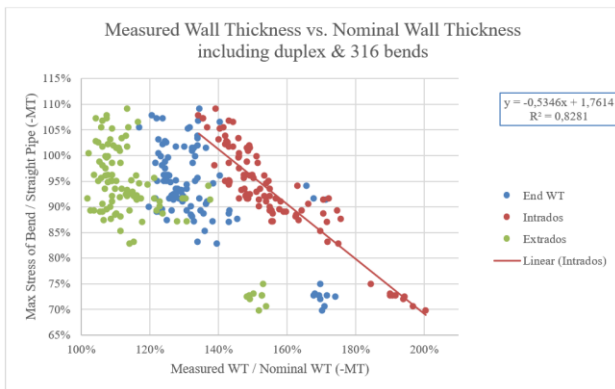
Figure 18. Result comparison of all three locations.

Calibration of the ultrasonic device in between each size, Temperature difference for the material expanding and contracting though this is minor, the material and resistance against rays, curvature which made measuring difficult and the use of zalo are all variables that may have had an effect on the results in this article.

4.3 Analyzing the Measurements

Early in this project, it was very clear the pipe bends do not have a uniform thickness. A simplification on the measurement locations, like the use of symmetry, would not accurately represent any given location between the bend end to the intrados/extrados.

There was a significant spread of how much thicker the bend was in comparison to the nominal pipe standard. This was especially noticeable in the samples that were of the same size but had multiple heat numbers and manufacturers.

In general, the results do not deviate from the expectations, but some variations were noticed. The NPS 4 SCH 80 is the one component size that is not in correspondence with the ASME B16.9 Standard based on the 6 point measurement. Detailed models of the NPS 4 SCH 80 could allow for better evaluation of the pipe. The simplified analysis of this pipe size is the only case where the bend was weaker than the corresponding straight pipe, which does not confirm ASME B16.9 Section 2.1. It is worth noting that the manufacturer of this particular batch is the same as the one with excessive wall thickness on the NPS 6 SCH 80.

With regards to the FEA model comparisons, the 40 point measurement shows a peak stress of over 500 MPa whereas the 20 point measurement is under the peak stress at 484 MPa. This implies a loss in between the modeling that is not covered by a standard spline fit in SOLIDWORKS model.

5 Conclusion

Based on the statistical analysis of the 81 measured bends, the lower bound values of the intrados with a 99.9% CI is 52.1% larger than the nominal wall thickness (-MT). The extrados is 12.4% larger than the nominal wall thickness (-MT), or nearly equal to the nominal wall thickness.

The database is still incomplete and should be expanded upon to be able to recognize any trends and map critical dimensions. There are 2 sizes (NPS 4 SCH 80 & NPS 6 SCH 80) that should be further investigated. There is insufficient data to confirm section 2.1 of ASME B16.9.

The most important comparison is the physical pressure test with the mathematical model which will verify if the modeling is good enough for further use, or if the technique needs adjustments.

The application of pipe bends in the subsea industry is crucial components and if they aren't close enough it can lead to costly consequences. Further research into the critical dimensions and deviation within the industry is recommended.

References

American Society of Mechanical Engineers. ASME Code for Process Piping, ASME B31.3. 2002.

American Society of Mechanical Engineers. Pipe Flanges and Flanged Fittings, ASME B16.9. 2012.

Det Norske Varitas. DNVGL-RP-F112.DNV, 2018.

Inspectioneering. Overview of Hydrogen Induced Cracking (HIC). 2020 [Online], Available: https://inspectioneering.com. [Accessed 20 April 2020]

CFD Simulations of Open Drainage System for Offshore Applications (Caisson) using OpenFOAM

Zidane Moujahidi   Joachim Lundberg

PEM-department, University of South-Eastern Norway, Norway, Joachim.lundberg@usn.no

Abstract

Flow assurance is of paramount importance in the oil and gas industry. In oil and gas platforms, caissons are used to discharge effluents at high volume flow rates. This latter can cause detrimental effects if they become dropped hazard objects, as they can damage subsea modules or gas pipelines. In the present work, the forces applied to a caisson due to the discharged seawater are assessed using multiphaseEulerFoam.

Keywords: caisson, OpenFOAM, VOF model, multuphaseEulerFoam

1 Background

Flow assurance is critical in the oil and gas industry since the flow of hydrocarbons, and other fluids should be transported successfully to ensure that the oil platform performs optimally. Interruptions of production due to any damages can cause significant financial losses.

Furthermore, in an oil platform, drainage systems are extensively present, and they can be either closed drainage systems or open drainage systems depending on the type of application. For instance, closed drainage systems, which are vented to flare, are used to collect gaseous oil and contaminated water while open drainage systems are vented to the atmosphere, and they are mainly used to dispose of seawater and non-contaminated water. Other applications of this latter include washdown water and rainwater plus process effluents and firewater (BP, 1992).

BP Group, a UK based oil and gas company, recommends taking several factors when designing or assessing open drainage systems. These factors include peak flows, hydraulic capacity, line sizing, as well as the degree of surcharge that can be tolerated (BP, 1992). In several circumstances, it is challenging to inspect a caisson, which results in inadequate assessment and little information about the state of the caisson or other equipment (BP, 2019). Nonetheless, this can be resolved by performing multiphase flow analyses to determine if there are any future potential risks.

In the context of the present study, an open drainage system i.e. a seawater caisson, is studied (Figure 1). These latter are vented tubular structures ranging over several floors designed to receive excess process streams. Also, they are used to discharge effluents continuously and intermittently with high momentum, and the flows often enter tangentially into the caisson. Noteworthy, these flows can cause detrimental effects such as exceeding the design pressure of one of the connecting supports or generating mechanical vibrations, which can, in turn, cause fatigue. Existing research on the field of multiphase flows in caissons are, to the author's knowledge, non-existing. However, some work or rotating annular flow has been investigated over the past years by Liu and Bai (2019) and Liu et al. (2020). Both studies was performed on much lower flow rates and equipment sizes than current study and the purpose of the study was to investigate the swirling flow itself rather than the forces to the surroundings.

Mohmmed et al. (2020) performed both numerical and experimental study on fluid-structure interactions from slugging multiphase flow in a horizontal pipe. The conclusion from the work was that the aeration rate affected the accuracy of the numerical models.

Stressman Engineering AS and other suppliers from the oil and gas sector are concerned about minimizing the cost and weight of the equipment, maintaining safe operation, and low maintenance. There are two cases to be addressed, in both, it is desired to quantify internal forces applied to the walls of the seawater caisson in order to determine if there are significant unbalanced forces that may cause vibrations or fatigue.

Figure 1. Seawater caisson including a support structure and membrane located in the right-hand side (White Rose, 2015). The structure is tilted horizontally in the figure.

2 Materials and Methods

The first case corresponds to the maximum single continuous seawater flow during normal operations while in the second case an additional seawater flow is added, thereby flow rates in the seawater caisson are increased. Also, the cases investigated using SpaceClaim as a CAD software, OpenFOAM as a CFD simulation tool, and Paraview as a postprocessing tool. Regarding the adopted methodology, the initial solid geometry was simplified by removing the nozzles or inlets in which no flow occurred. Afterward, the fluid domain was extracted from solid geometry. Then, polygon grids were generated for each of the fluid domains. The two cases were simulated using multiphaseEulerFoam. The Eulerian approach was expected to result in more realistic flow patterns and a better depiction of phase fractions in the fluid domain.

No experimental data or comparable CFD studies was available, so the results were checked scrutinized, compared, and screened qualitatively and quantitatively.

3 Fluid Domains and Boundary Conditions

After simplifying the initial geometry, the fluid domain is extracted from the solid geometry of the seawater caisson using SpaceClaim software. The boundary conditions of the fluid domain are given in Table 1.

The total length of a seawater caisson is 24.34 [m], and constituted two parts. The first part, zone 1, is 21.45 [m] long, and the second part, zone 2, is 2.89 [m] long, which corresponds to the transition zone. Figure 2 illustrates the two fluid domains.
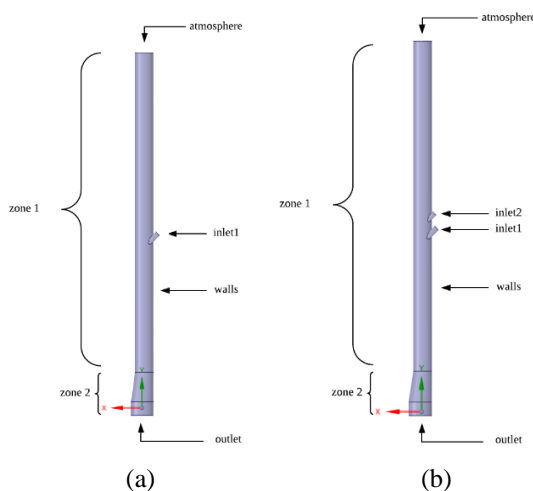
Figure 2. (a) First fluid domain; (b) second fluid domain; boundary names; and the two zones.

The inlet is inclined by 45°, and it has a length of 0.53 [m]. Then, it is connected to zone 1, and it is located approximately 12.00 [m] relatively from the outlet. Noteworthy, the seawater enters the seawater caisson tangentially and, more specifically, through zone 1. The

corresponding diameters of inlet1, outlet, and the atmosphere are shown in Table 1.

Table 1. Dimensions of the patches.

Boundary condition	Diameter [m]
inlet1	0.36
inlet2	0.30
outlet	1.50
atmosphere	1.20

4 Mesh

Initially, tetrahedral meshes were generated, and then they were converted to polyhedral meshes with a feature angle of 70 using the utility polyDualMesh 70. Afterward, the mesh was exported as a .msh file in ASCII format.

According to Balafas (2014), one of the main advantages of polyhedral mesh over tetrahedrons is that polyhedral cells are bounded by adjacent cells, which makes the computation of gradients significantly better. Moreover, polyhedral cells are less prone to stretching compared to the other cell types making them a good candidate for preserving good mesh qualities, thereby reducing numerical diffusion and improving the accuracy of the results at a lower cell count (Balafas, 2014). Table 2 shows the mesh metrics.

Table 2. Mesh metrics of the fluid domains 1 and 2.

Metrics	Fluid domain 1	Fluid domain 2
Total number of cells	599344	990071
Number of inflation layers	10	10
Max. skewness	0.590	0.550
Avg. skewness	0.012	0.013
Max. non-orthogonality	0.600	0.540

Since two flows were involved in the second case study, it was expected that large gradients might occur, especially when the two flows interact at the wall of the caisson. A refined mesh was used to depict the gradients as shown in Figure 3.

Also, in the present work, the wall functions approach is used with $30<y+<200$. The employed wall functions are mainly nutkWallFunction, kqRWallFunction, and epsilonWallFunction.

Moreover, since the CFD results were verified and were found to be close to the hand calculations, no further mesh refinement was performed. In other words, the mesh independence study was not carried out.
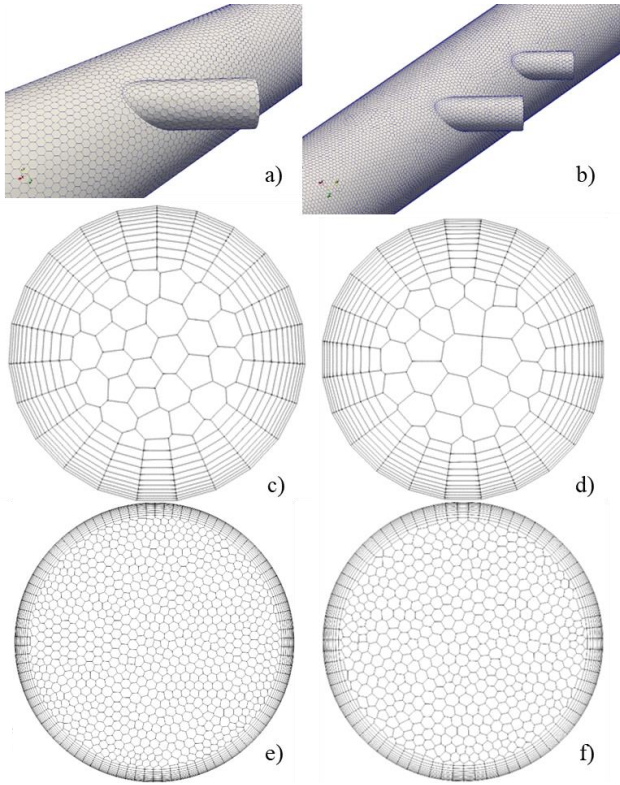
Figure 3. Meshing of the two fluid domains: (a) and (b) walls; (c) inlet1; (d) inlet2; (e) outlet; and (f) atmosphere.

5   MultiphaseEulerFoam

MultiphaseEulerFoam is a multiphase flow solver that can be used to solve for two or more compressible phases. It is implemented based on an Eulerian approach, and it can be coupled with the VOF interface sharpening method and turbulent models, such as RANS and LES models.

In the Eulerian approach, the phases are treated as interpenetrating continua, and the momentum equations are solved for each fluid separately; hence each fluid has its velocity field. However, the two phases have the same pressure field. Then, the momentum equations are coupled through the pressure filed, and the interface exchange momentum forces. The mass equations per phase are

$$\frac{\partial \rho_q \alpha_q}{\partial t} + \nabla \cdot \left(\rho_q \alpha_q \boldsymbol{u_q} \right) = 0 \qquad (1)$$

Where

q	seawater or air
α	Phase fraction
$\boldsymbol{u_q}$	Velocity field
ρ_q	Density of q

Then, the momentum equations for each of the phases are

$$\frac{\partial \rho_q \alpha_q \boldsymbol{u_q}}{\partial t} + \nabla \cdot \left(\rho_q \alpha_q \boldsymbol{u_q} \boldsymbol{u_q} \right) = \\ -\alpha_q \nabla p + \nabla \cdot \left(\alpha_q \tau_q \right) + \rho_q \alpha_q \boldsymbol{g} + \boldsymbol{M_q} + \boldsymbol{F_\sigma} \qquad (2)$$

Where

p	Static pressure
τ_q	Stress tensor
M_q	Momentum transfer
F_σ	Surface tension force

The stress tensor is expressed as

$$\tau_q = \left(\mu_q + \mu_t \right) \left(\nabla \boldsymbol{u_q} + \left(\nabla \boldsymbol{u_q} \right)^T - \frac{2}{3} \nabla \cdot \boldsymbol{u_q} \boldsymbol{I} \right) \qquad (3)$$

Where

I	Unit second-order tensor
μ_q	Viscosity
μ_t	Eddy viscosity

The transport equation of the phase fraction is expressed as

$$\frac{\partial \alpha_q}{\partial t} + \boldsymbol{u_q} \cdot \nabla \alpha_q + \nabla \cdot \left(\boldsymbol{u_r} \, \alpha_q \left(1 - \alpha_q \right) \right) = 0 \qquad (4)$$

The third term in Equation 4 represents the compression term, and it is used instead of the interface-capturing algorithms in multiphaseEulerFoam. According to Wardle and Weller (2013), the divergence of the compression velocity $\boldsymbol{u_r}$ guarantees the conservation of the volume fractions, while the term $\alpha(1 - \alpha)$ restricts the effect of the interface sharpening only at the interface. Moreover, the compression term is based on the maximum velocity magnitude in the transition region, and it is expressed as

$$\boldsymbol{u_r} = \min(C_\alpha |\boldsymbol{u}|, \max(|\boldsymbol{u}|)) \vec{\boldsymbol{n}} \qquad (5)$$

And

$$\vec{\boldsymbol{n}} = \frac{\nabla \alpha}{|\nabla \alpha|} \qquad (6)$$

Where

u_r	Compression velocity
C_α	Compression factor
$\vec{\boldsymbol{n}}$	Normal vector at the interface

Moreover, a continuity equation for the mixture is needed to solve for the implicit pressure equation, thus

$$\nabla \cdot \boldsymbol{u} = 0 \qquad (7)$$

$$\boldsymbol{u} = \alpha_w \boldsymbol{u_w} + \alpha_a \boldsymbol{u_a} \qquad (8)$$

6   Simulation Setups

In both case studies, the pressure at the inlets was set to fixedFluxPressure, and the velocities at the inlets specified using fixedValue. The velocity at inlet1 was specified to 6.13 $[m \, s^{-1}]$ and inlet2 to 3.00 $[m \, s^{-1}]$.

For the outlet patch, prghPressure boundary condition with a static reference pressure of 101325 [Pa] was used along with the boundary condition pressureInletOutletVelocity. It switches between zeroGradient and fixedValue. zeroGadient was applied in flow directions out of the domain, while fixedValue was applied for backflow into the fluid domain.

Concerning the atmosphere patch, a common way to specify the boundary condition at this type of patch is by using totalPressure for the pressure and the

pressureInletOutletVelocity for the velocity. totalPressure Boundary condition behaves in two ways depending on the direction of the flow. On the one hand, when there is an outflow, the pressure is set to the reference pressure, i.e., atmospheric pressure. On the other hand, when there is an inflow into the fluid domain, the pressure was computed using Equation 9 (Greenshields, 2018).

$$p = p_{ref} - \frac{1}{2}|\boldsymbol{u}^2|$$ (9)

About the initialization of the turbulence properties, i.e., k and ϵ, at the patches were calculated and specified using Equation 10 to Equation 13. First, the Reynolds number was calculated using the inlet pipe diameter and seawater thermophysical properties.

$$Re = \frac{\rho_w u_{inlet} D_{pipe}}{\mu_w}$$ (10)

Secondly, according to Russo and Basse (2016), the turbulence intensity was approximated using

$$I = 0.140 Re^{-0.0790}$$ (11)

Other equations can be used to estimate the turbulence intensity. However, it was noticed that the last equation leads to a smoother solution process. The turbulence kinetic energy and the rate of dissipation at the inlets are estimated using

$$k = \frac{3}{2}(\boldsymbol{u}_{inlet}I)^2$$ (12)

$$\varepsilon = C_\mu^{\frac{3}{4}} \frac{k^{\frac{3}{2}}}{l}, \qquad l = 0.07\, D_{inlet}$$ (13)

It is important to note that the internal field was initialized using the lowest values obtained from the calculation of the turbulence properties at the inlets, in this case, inlet2. This procedure helps the solvers to compute the governing equations faster. The turbulence properties are summarized in Table 3

Table 3. The turbulence properties at the patches.

Patches	$k\ [m^2 s^{-2}]$	$\varepsilon\ [m^2 s^{-1}]$
inlet1	0.140	0.340
inlet2	0.033	0.047
internalField	0.033	0.047

Furthermore, specifying the fluid properties is crucial since it predetermines the behavior of multiphase systems. It was assumed that both seawater and air are incompressible and isothermal fluids. Besides, the thermophysical properties of seawater and air are taken at 5°C and standard atmospheric pressure. The fluids properties are shown in Table 4

Table 4. Fluid properties.

Fluids	$\rho\ [kg\ m^{-3}]$	$\upsilon\ [m^2\ s^{-1}]$
Seawater	1028	1.56·10⁻⁶
Air	1.26	1.37·10⁻⁵

multiphaseEulerFoam prompts the user to enter the diameter of the dispersed phase and the continuous

phase. Due to the lack of data, the droplet size and bubble size are assumed to 1.00·10⁻⁴ [m] 3.00·10⁻³ [m], respectively.

Regarding the time stepping in the present analysis, adjustable time stepping was used and the average throughout the simulation time was found to be 0.001[s].

7 Results and Discussion

The upper halves of walls of the caisson and more specifically, above the upper inlets are not considered since the forces are expected to be insignificant. Figure 5 shows the divided walls, and the alternation between orange and white colors is for illustration purposes. The walls of the first and second fluid domains are divided into 13 slices and 14 slices, respectively. The slices are 1.00 [m], except for the last slices located at the bottom of the caisson, which is 0.36 [m].

Only the most descriptive slices are discussed in this paper. For more information about the forces applied to each slice are provided the work by Zidane (2020).

Additionally, in subsection 9.1 and subsection 9.2, the force plots show the net forces and the forces in the three. The black curve represents the force components along the X-axis (axially with the inlets), the red curve represents the force components along the Y-axis (axially with the caisson), and the blue curve represents the force components along Z-axis (perpendicular to the inlets). The green curve represents the magnitude of the forces. In the phase fraction contours, the location of the maximum net force is highlighted with a yellow dot.

(a) (b)
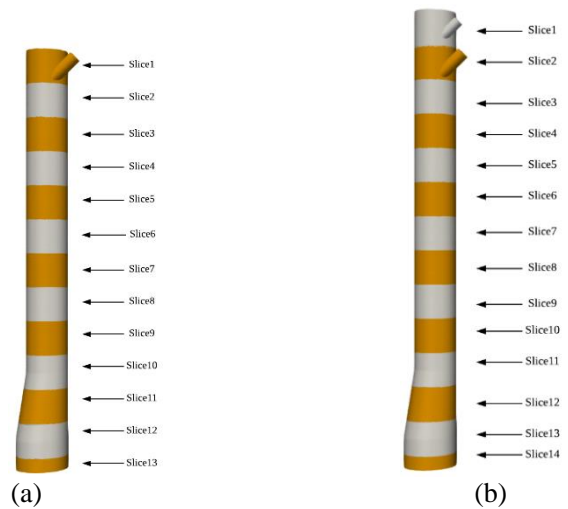
Figure 5. (a) Walls of the caisson segmented into 13 slices (first case study); and (b) walls of the caisson segmented into 14 slices (second case study).

7.1 Evolution of peak forces – Case 1

The seawater flow accelerates due to the 45 ° inclination of the inlet pipe and the free fall between the exit of the pipe inlet and the wall of the large tubular structure. Besides, the seawater flow reaches 11.56 [$m\ s^{-1}$] when

it contacts the curved wall of the caisson. Figure 6 illustrates the net force that increases abruptly reaching a peak value of 14550 [N] at 0.30 [s], however, the net force decreases to 11000 [N] and remains at this value with minor fluctuations, more specifically after 2.40 [s]. Figure 7 shows the concentration of the net force in slice2 at 0.30 [s].

After slice2, it is observed that the forces applied to the walls of the caisson decrease, which implies that the energy of the seawater flow diffuses as the flow progresses downward. The decrease in force occurs mainly from slice4 to slice9, and it is accompanied by fluctuations that last for a certain period; however, the force fluctuations vanish thereafter.
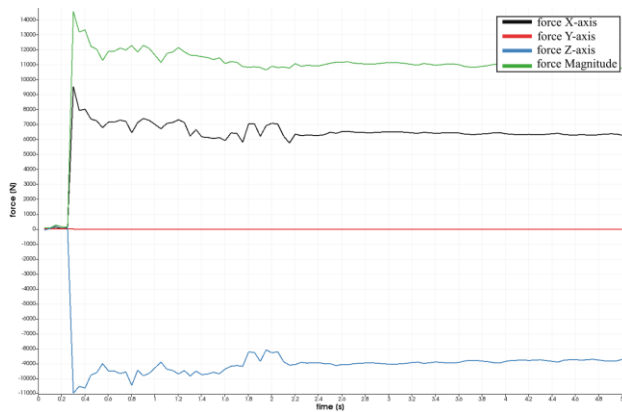

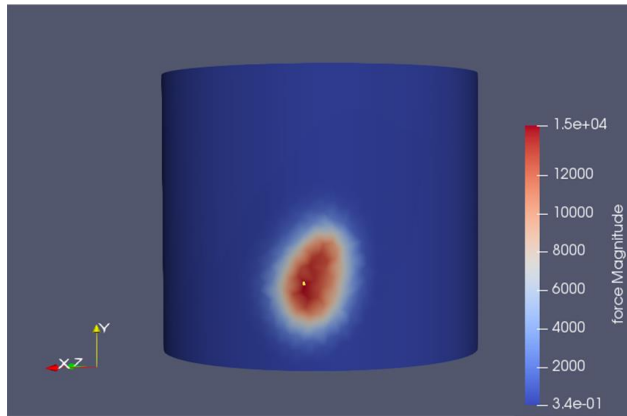**Figure 6.** Slice2 - located at (0, -1.3, 0).

Figure 7. Force magnitude contour at 0.3 [s] in slice2-located at (0, -1.3, 0).

Following slice11, the forces become significant again, for example in slice13, which is at the outlet of the caisson, the forces increase in a fluctuating way and become slightly stable at 13000 [N] starting from 4.40 [N], and the load imbalance is primarily due to the X component of the net force, as shown in Figure 8.

A critical remark is that the force imbalance in slice13 is due to the accumulation of seawater along a line due to the swirling effect, i.e., the swirling streak. Figure 9 proves that the maximum net force coincides with the location of the swirling streak.
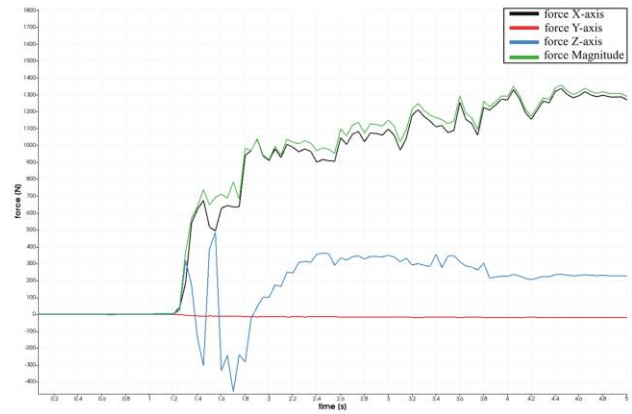
Figure 8. Slice13 - located at (0, -12.3,0).
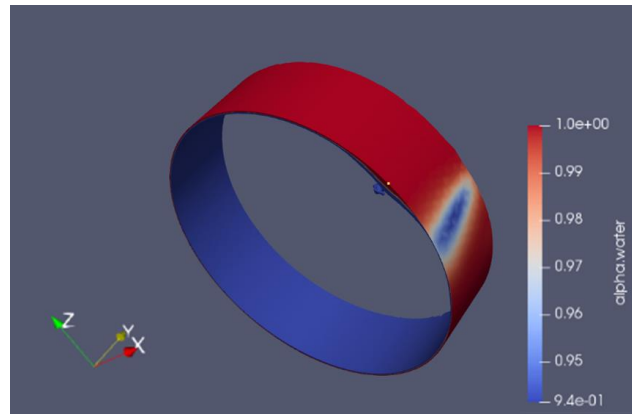
Figure 9. Seawater phase fraction contour at 5.00 [s] in slice13 - located at (0, -12.3, 0).

7.2 Evolution of peak forces – Case 2

In slice3, the net peak force is 15980 [N] at 0.40 [s], which is higher than the value found in the first case study (slice2). More interestingly, the value of the net force when the flow stabilizes is 7700 [N] which is lower than the value found in the first case of 11000 [N]. Besides, the force imbalances are due to the X and Z components of the net force. The evolution of the forces in slice3 is shown in Figure 10.

Another interesting finding is that the momentum of the seawater at inlet1 transports the momentum of the seawater flow at inlet2 down the caisson causing the forces to be more spread out in the slices below slice2. Also, it is noticed that the forces down the caisson display more fluctuations compared to the first case study, which is caused by the presence of the two seawater flows, i.e., instabilities, and these observations are best described using the evolution of the forces at slice5. Figure 11 shows that the maximum net force is 1550 [N], and it occurs at 1.40 [s] and the X component of the force appears to be the main contributor to the net force. Additionally, the forces exhibit large fluctuations, for example, at 2.65 [s], the fluctuation of the force along the X-axis is about 2000 [N].
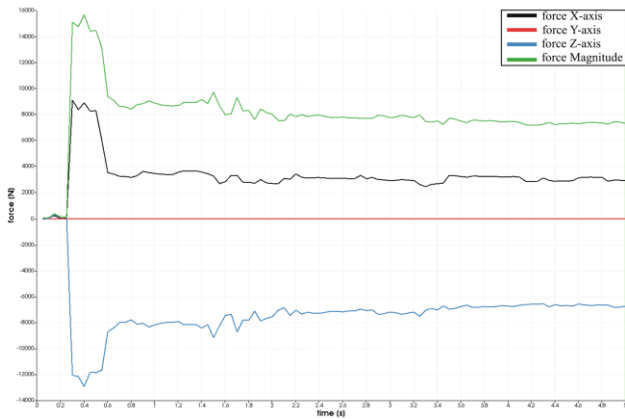
Figure 10. Slice3- located at (0, -1.3, 0).
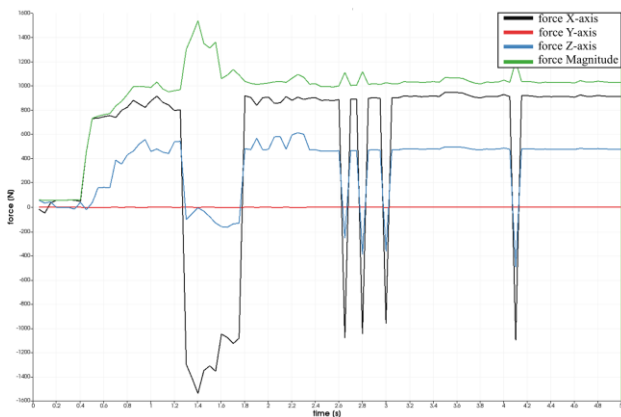
Figure 11. Slice5 - located at (0, -3.3, 0).

Interestingly, Figure 12 illustrates that the most significant force in slice5 coincides with the leading wave of the swirling flow. Also, all the large fluctuations of forces in the X and Z axes are in the same position as swirling steak; however, when the large fluctuations disappear, and the position of the forces move to another position.
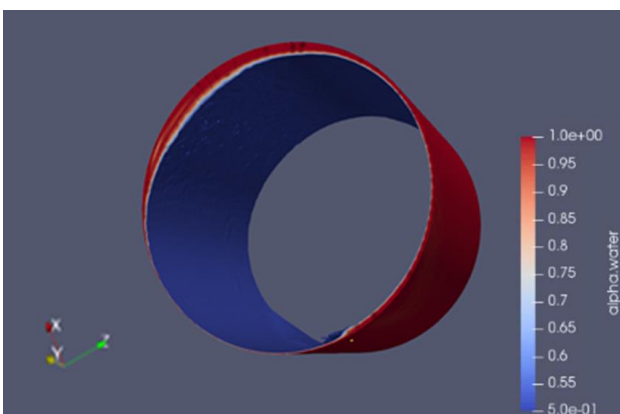
Figure 12. Seawater phase fraction contour at 1.40 [s] in slice5 - located at (0, -3.3, 0).
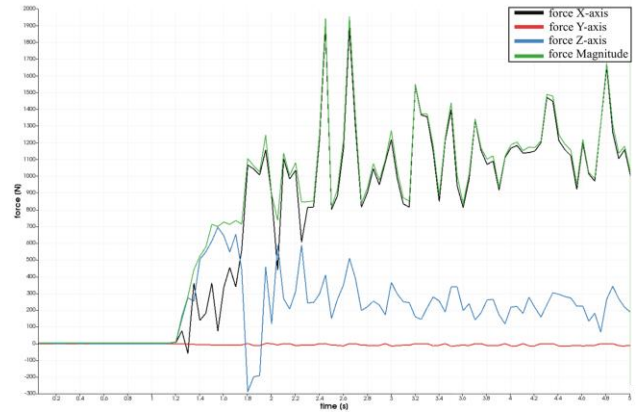
Figure 13. Slice14- located at (0, -12.3, 0).

It must be emphasized that the forces display similar trends as in the first case study except that the forces are higher and more fluctuating, which is caused by the presence of two seawater flows and a higher volume flow rate entering the caisson. For example, Figure 13 shows that the large fluctuations in slice14 occur primarily due to the force component in the X-direction.

Figure 14 shows that the location of the maximum force coincides with the swirling streak location, and it is the main cause of the force fluctuations in slice14.
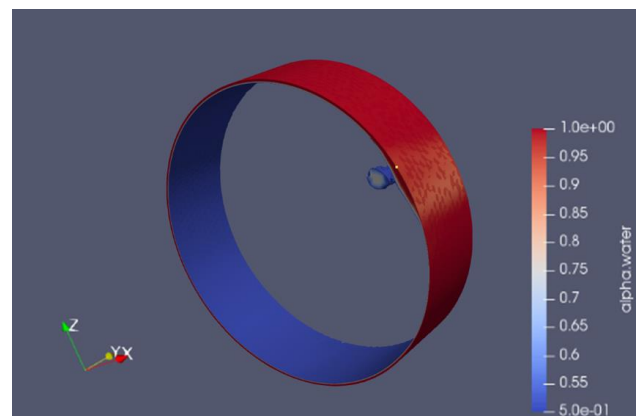
Figure 14. Seawater phase fraction contour at 5.00 [s] in slice14 - located at (0, -12.3, 0).

7.3 Sanity check

When the seawater flow is ejected from inlet1, its velocity increases up to 13.62 $[m\,s^{-1}]$ right after impacting the wall, thus the flow gains momentum. To calculate the force induced by the flow, a simple hand calculation is used

$$\boldsymbol{F_w} = \dot{m}_w \boldsymbol{u_w} = \rho_w A_{inlet1} \boldsymbol{u_w^2} \qquad (14)$$

$$1028 \cdot \left(\pi \cdot \left(\frac{0.36}{2} \right)^2 \cdot (13.62 \cdot \cos(45))^2 \right)$$
$$= 9705\ [N]$$

This value is close to the X-component of the net force in slice2 (0, -1.3 ,0), as shown in Figure 6. Besides, from the latter figure, it is observed that the force in Z-component is slightly higher than the force in the X-

component, and this can be explained by the curvature of the walls of the caisson which causes the flow to accelerate in the Z-direction more than in the X-direction. Also, it is sensical the energy of the seawater flow diffuses as it progresses downward, except in the transition zone where it increases because of seawater accumulation and disturbances due to the change in geometry.

Additionally, from a qualitative perspective, the flows in the caisson behavior appear to be physical, i.e., vertical downward annular flows and the presence of swirling streak as found by Liu et al. (2020).

8 Conclusions

From the CFD simulation results of the two cases, the following conclusions are drawn:

- The magnitude of the forces and their distribution in the caisson are strongly dependent on the volume flow rates and the number of the inlets.

- The main forces are acting in the radial direction whereas the forces in the axial direction are nonexistent in all the parts of the caisson and throughout all the simulation times.

- In the first case study, the largest forces are located where the seawater interacts with the walls of the caisson for the first time, namely, in slice2 and slice3 and slice13. Then, as the flow of seawater progresses, the forces decrease in magnitude, implying that the energy of the seawater is diffused. However, when the seawater reaches the transition zone, the forces further increase and display more fluctuations as a result of the increased momentum of the seawater flow, flow accumulation, and flow disturbances.

- In the second case, the forces have similar trends as in the first case, but they are larger and more fluctuating, and they are more distributed throughout the lower section of the caisson. This happens because the seawater flow from inlet2 interferes with the seawater flow from inlet1, causing more instabilities in the seawater flow. Also, this interference causes the propagation of seawater flow momentum in the downward direction, hence the different force distributions between the first and second cases

- In both cases, these forces exhibit fluctuations; however, they vanish afterward.

- The mass flow rates through the caisson increase exponentially and reach pseudo-steady states after the flow establishes.

- It is noticed that air is sucked from the atmospheric boundary due to the downward momentum of the seawater flow.

- The first and third cases show good agreement with the hand calculations.

- All the simulations show reasonable flow patterns which are in good agreement with flow phenomena mentioned in the literature review.

- Since large, imbalanced, and fluctuating forces are found in the case studies, it is recommended to use the CFD results in present work to perform FEA analyses to quantify the resulting vibrations and fatigues.

References

G. Balafas. *Polyhedral Mesh Generation For CFD - Analysis Of Complex Structures*. Master's thesis, Technical University of Munich, Munich, Germany, 2014.

BP. *Azeri Central East Project: Environmental and Social Impact Assessment, Non-Technical Summary*, BP, 2019.

BP. *Guide to the Selection, Arrangement and Specification of Offshore Platform Drainage*. Engineering Practices Group, BP International Limited, Middlesex, United Kingdom, 1992.

C. Greenshields. *Openfoam V6 User Guide: 5.2 Boundaries*. CFD Direct, Retrieved from https://cfd.direct/openfoam/user-guide/v6-boundaries. July 10th 2018.

L. Liu and B. Bai. A mechanistic model for the prediction of swirling annular flow pattern transition. *Chemical Engineering Science*, 199:405-416, 2019. doi: 10.1016/j.ces.2019.01.039.

W. Liu, X. Lv and B. Bai. Axial development of air–water annular flow with swirl in a vertical pipe. *International Journal of Multiphase Flow*, 124, 2020. doi: 10.1016/j.ijmultiphaseflow.2019.103165.

A. O. Mohmmed, H. H. Al-Kayiem. A. B. Osman and O. Sabir. One-way coupled fluid–structure interaction of gas–liquid slug flow in a horizontal pipe: Experiments and simulations. *Journal of Fluids and Structures*, 97, 2020. doi: 10.1016/j.jfluidstructs.2020.103083.

F. Russo and N. T. Basse. Scaling of turbulence intensity for low-speed flow in smooth pipes. *Flow Measurement and Instrumentation,* 52:101-114, 2016. doi: 10.1016/j.flowmeasinst.2016.09.012.

K. E. Wardle and H. G. Weller. Hybrid Multiphase CFD Solver for Coupled Dispersed/Segregated Flows in Liquid-Liquid Extraction. *International Journal of Chemical Engineering*, 2013:1-13, 2013. doi: 10.1155/2013/128936.

White Rose. *K30 Storage Process Description, Technical: Storage*. UK CCS Commercialisation Programme, United Kingdom, December, 2015.

M. Zidane. *CFD Simulations of Open Drainage System for Offshore Applications (Caisson) Using Openfoam*. Master's thesis, University of South-Eastern Norway, Porsgrunn, Norway, 2020.

Model Based Early Kick/Loss Detection and Attenuation with Topside Sensing in Managed Pressure Drilling

Asanthi Jinasena[1]   Haavard Holta[2]   Morten Hansen Jondahl[3]   Prasanna Welahettige[4]   Ole Morten Aamo[2]   Roshan Sharma[3]   Håkon Viumdal[3]   Bernt Lie[3]

[1]Department of Energy and Process Engineering, Norwegian University of Science and Technology, Norway, `asanthi.jinasena@ntnu.no`
[2]Department of Engineering Cybernetics, Norwegian University of Science and Technology, Norway, `{haavard.holta,aamo}@ntnu.no`
[3]Department of Electrical Engineering, Information Technology and Cybernetics, University of South-Eastern Norway, Norway, `morthans1@gmail.com`, `{roshan.sharma,Hakon.Viumdal,Bernt.Lie}@usn.no`
[4]Department of Process, Energy and Environmental Technology, University of South-Eastern Norway, Norway, `prasanna.welahetti@gmail.com`

Abstract

Early kick/loss detection is a crucial part of safe well control, and it plays a major role in the reduction of risk and non-productive time in drilling. In conventional drilling, topside sensing is used for early kick/loss detection. Recently, Venturi flowmeter based online return flow estimation has been introduced for this purpose by the authors. In managed pressure drilling, both topside sensing and bottomside sensing can be used for kick/loss detection. Therefore, a topside return flow estimator with a bottomside well pressure and flow estimator is combined to provide a complete kick/loss detection and estimation scheme for managed pressure drilling systems. This allows improved kick/loss detection. In addition, a closed-loop kick/loss attenuation controller is used to illustrate the estimation scheme.

Keywords: kick loss detection, managed pressure drilling (MPD), return flow, adaptive control, UKF

1 Introduction

1.1 Background

Oil and gas drilling is done by penetrating a rotating drill bit into the rock formation, creating a wellbore. The formation is a high pressure and temperature environment. Therefore, a drilling fluid known as mud is continuously circulated through the wellbore. This circulation process is usually divided into two parts: bottomside (wellbore) and topside (section on the surface).

The bottomhole pressure (BHP) is maintained within a window for safe operation. If BHP is lower than the formation fluid pressure, formation fluid may enter the wellbore. This phenomenon is commonly referred to as a 'kick' which could result in a catastrophic blowout if not controlled properly. If BHP is higher than the fracture pressure of the formation, the drilling mud may seep into the formation, which is known as a 'loss'.

Early kick/loss detection is a crucial part of safe well control, and it plays a major role in the reduction of risk and non-productive time in drilling. In conventional drilling, this is primarily achieved by using the topside data, mainly the return mud flow measurement, and the volume gain in the mud pit. For the return flow, cost-effective, accurate and online sensors are needed in this regard. For applications with narrow pressure margins, a control choke and a back-pressure pump are used to control the BHP fast and accurately. This is known as managed pressure drilling (MPD).

The topside sensing includes among other measurements: return flow measurements, mud pit gains, and other respective rheological properties of the drilling mud that is essential for normal drilling operations. However, the rheological measurements are often offline, manual measurements with low frequencies and not eligible for automation purposes.

1.2 Previous Work on Topside Sensing

A Venturi flowmeter based online return flow estimation has been studied recently. The fluid level changes in the open channel is measured non–intrusively and used to model the fluid flow rate. Both mechanistic and data-driven models are used for this purpose (Berg et al., 2015; Chhantyal et al., 2018; Jinasena et al., 2018; Welahettige, 2019; Jinasena, 2019; Jondahl, 2020; Berg, 2020). Further, these fluid level characterizations are used as inputs to different data-driven models for the estimation of the drilling fluid rheological properties (Chhantyal et al., 2016). Moreover, some of the fluid rheological properties which are essential for the drilling operation (mainly viscosity and density) are estimated using ultrasonic waves in a stationary medium (Jondahl and Viumdal, 2018, 2019; Jondahl, 2020). The estimations were done by using different data-driven models, and the estimation errors were within the NORSOK standards accuracy of 2% (Jondahl, 2020).

Further, the viscosity and the flow behaviour of the drilling mud has been studied extensively using 3D computational fluid dynamics (CFD) simulations (Welahettige, 2019). This has been beneficial to the different models that can be used for return flow rate estimation. High-resolution numerical schemes such as the flux-limited-centered scheme has been applied to solve a non-Newtonian model for open channel. This scheme is well-balanced, positivity-preserving and has a high accuracy, and a good resolution for discontinuities. The developed model was tested with hydraulic jumps propagation in open Venturi channels. Moreover, the effect of drill cuttings on the return flow measurements have also been studied (Welahettige et al., 2019) and a multi-fluid volume of fluid model has been used for the CFD analyses.

These studies as a whole, improve the topside models and the estimations. Further, these studies complement the topside sensing and the early kick detection in general.

In this paper, we combine a topside return flow estimator with a bottomside well pressure and flow estimator to provide a complete kick/loss detection and estimation scheme for MPD systems. In addition, to illustrate the estimation scheme, a closed loop kick/loss attenuation controller is used.

The paper is organized as follows. The complete system, including the mathematical models, estimation and control methods, are presented in Section 2. This is followed by the simulation set-up, detailed results and discussion in Section 3. Finally, the conclusions drawn from the results and discussion are summarized in Section 4. Further, a more detailed mathematical overview of the estimator schemes can be found in the Appendix A.

2 System Description

The block diagram of the considered complete flow loop of the MPD system is shown in Figure 1. The main process steps of the entire flow loop which are modeled are shown here with the notations.

2.1 Mathematical Models

Both the flow dynamics in the annulus and the top-side Venturi channel are modelled as distributed systems of hyperbolic PDEs, while the drill string is considered as a known input-output system. In the annulus, the pressure $p(z,t)$ and flow rate $q(z,t)$ are modelled by the well known *water hammer equations* for compressible, single-phase flows with low Mach numbers (Ghidaoui et al., 2005), while the wetted cross-sectional area $A_c(x,t)$ and flow rate $q_l(x,t)$ in the Venturi channel are modelled by the Saint-Venant, shallow water equations (Chow, 1959; Chaudhry, 2008).

The water hammer equations modelling the well flow have the form,

$$\frac{\partial p(z,t)}{\partial t} = \frac{\beta}{A_a}\frac{\partial q(z,t)}{\partial z}, \qquad (1)$$
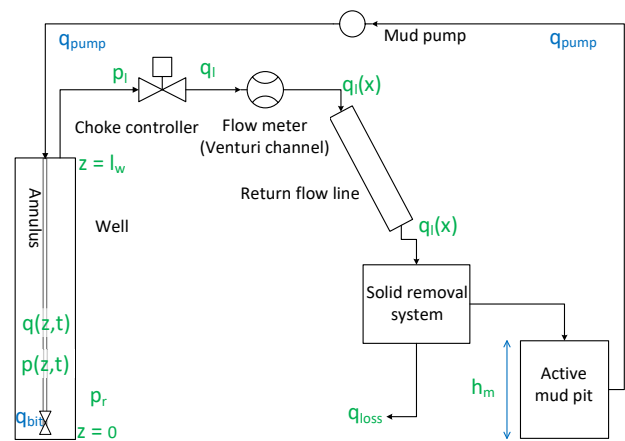
Figure 1. The block diagram of the flow loop of a MPD system.

$$\frac{\partial q(z,t)}{\partial t} = -\frac{A_a}{\rho}\frac{\partial p(z,t)}{\partial z} - \frac{F_a}{\rho}q(z,t) - A_a g. \qquad (2)$$

Here, $z \in [0, l_w]$ and $t \geq 0$ are the independent variables of space and time respectively, where $z = l_w$ is at the top of the well and l_w is the well depth. ρ is the density of the drilling mud, β is the bulk modulus of the mud, F_a is the friction factor in the annulus, A_a is the cross sectional area of the annulus and g is the acceleration of gravity.

The two boundary conditions are the topside pressure $p(l_w, t)$, and the net inflow $q(0, t)$ at the bottom of the well. The topside pressure $p_l(t)$ is related to the topside flow rate $q(l_w, t)$ through the choke equation,

$$q(l_w, t) = k_{\text{choke}}(t)\text{sign}(p_l(t) - p_0)\sqrt{|p_l(t) - p_0|}, \quad (3)$$

where p_0 is the atmospheric pressure and k_{choke} is the choke coefficient which can be used as an actuation. The net inflow is modelled as a simple linear flow-pressure relationship as follows (Dake, 1998),

$$q(0,t) = J\left(p_r - p(0,t)\right) + q_{\text{bit}}, \qquad (4)$$

$$p(l_w, t) = p_l(t), \qquad (5)$$

where p_r is the pressure in the surrounding formation, q_{bit} is a known volumetric flow through the drill bit, J is the so-called *productivity-index*. Both the productivity index $J > 0$, and the formation pressure $p_r > 0$ are assumed unknown and must be estimated. The topside flow rate $q_l(t)$ is modeled using the topside model which is described next.

The topside flow rate $q_l(x,t)$ is measured by a Venturi channel flowmeter, which uses the fluid levels $h(x,t)$ to calculate the flow rate based on the 1–D shallow water equations (Chow, 1959; Chaudhry, 2008).

$$\frac{\partial A_c(x,h,t)}{\partial t} = -\frac{\partial q_l(x,t)}{\partial x} \qquad (6)$$

$$\frac{\partial q_l(x,t)}{\partial t} = -\frac{\partial}{\partial x}\left(\alpha\frac{q_l^2(x,t)}{A_c(x,h,t)} + gI_1\cos\phi\right) + gI_2$$
$$+ gA_c(x,h,t)\sin\phi - T_f \qquad (7)$$

Here, $A_c(x,h,t)$ is the wetted cross sectional area in the channel normal to the flow, $h(x,t)$ is the depth of flow, and $q_l(x,t)$ is the volumetric flow rate in the channel. $x \in [0, l_c]$ is the position along the channel, and l_c is the length of the channel. I_1, the first moment of area represents the hydrostatic pressure term and I_2 represents the pressure forces in the fluid volume, which occur from the longitudinal width and slope variations in the channel. ϕ is the channel bottom slope angle, α is known as the momentum correction coefficient or the Boussinesq coefficient and corresponds to the deviations of the local velocity over the mean velocity of the flow, and T_f is the non-Newtonian friction term.

Equations 6 and 7 are used for the Venturi channel with a trapezoidal cross section and a zero bottom slope angle. The boundary condition for the Venturi channel is the topside flow rate $q(l_w, t)$ of the choke equation (equation 3). The fluid levels produced by the model will be then taken as measurements for the estimation.

The same set of equations (6 and 7) is used to model the return flowline with a circular cross section and a 7° bottom slope angle. The flowline is connected to the Venturi by taking the Venturi outlet flow rate $q_{l_3}(t)$ as the boundary condition to the flowline model. More details on the Venturi channel model and the return flowline model can be found at (Jinasena et al., 2019; Pirir et al., 2018), respectively.

The active mud pit level h_m is modeled by the simplified mass balance equation as follows (Pirir et al., 2018),

$$\frac{dh_m}{dt} = \frac{1}{A_m}(q_l - q_{loss} - q_{pump}), \qquad (8)$$

where A_m is the cross sectional area of the active mud pit, q_{loss} is the fluid losses at solid removal equipment and q_{pump} is the mud flow in to the well which is measured by the mud pump. Here q_l is the outlet flow rate of the flowline.

The density of the drilling mud is assumed to be constant throughout the flow loop, and only a single phase (liquid) flow is considered with no gas or cuttings flow.

The return flow rate from the Venturi channel, and the mud pit level are estimated. Further, the topside model allows to estimate the non-Newtonian friction coefficient of the mud and the fluid losses from the solid removal equipment as well (Jinasena et al., 2019; Jinasena and Sharma, 2020). However, the friction coefficient and fluid losses are not estimated in this study as the focus is on the well dynamics.
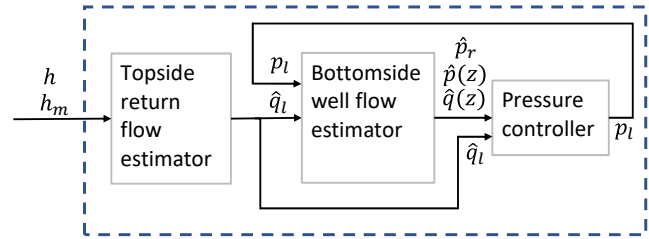
Figure 2. Estimators and controller configuration.

2.2 Estimation and Control Methods

The main objective of this study is to use this combined model to estimate the reservoir pressure and the BHP, which can be an input to the kick/loss attenuation system.

The system configuration is shown in Figure 2. The available measurements for the system are the fluid levels of the Venturi channel and the mud pit level. The topside estimator uses both of these measurements to estimate the return flow rate. The return flow estimates are used as an input to the bottomside estimator, where the annulus pressure and flow, and the reservoir pressure and productivity-index are estimated. All estimated states and parameters are then used to guide a closed-loop kick/loss attenuation system by adjusting the top-side MPD choke opening.

In this study, for the topside estimator, we use the estimator presented in (Jinasena et al., 2019) which is based on an unscented Kalman filter (UKF). For the bottomside estimator we use the adaptive observer design presented in (Holta et al., 2017).

2.2.1 UKF

The return flow rate, the non-Newtonian friction coefficient and the rate of fluid loss at the solids removal system can be estimated by the UKF (Jinasena and Sharma, 2020). One of the fluid levels of the Venturi channel can be used as an input, if preferred. However, in this study, two fluid levels are used as measurements and no input is used for the estimator. Further, only the return flow rate is estimated for simplicity. More details on the UKF can be found in (Jinasena et al., 2019; Jinasena and Sharma, 2020) and a brief explanation is given in Appendix A.2.

2.2.2 Adaptive Observer

The distributed pressure $p(z,t)$ and flow $q(z,t)$ in the annulus, are estimated using a distributed-state observer which is derived using the so-called *infinite-dimensional backstepping* design. The unknown reservoir pressure p_r and the productivity index J are estimated by an adaptive law based on a linear parametric model. Both the topside choke pressure and the estimated return flow are considered as external inputs to the observer. All necessary design elements needed to implement the adaptive observer are provided in Appendix A.1. More details on the observer design, including stability proofs can be found in (Holta et al., 2017).
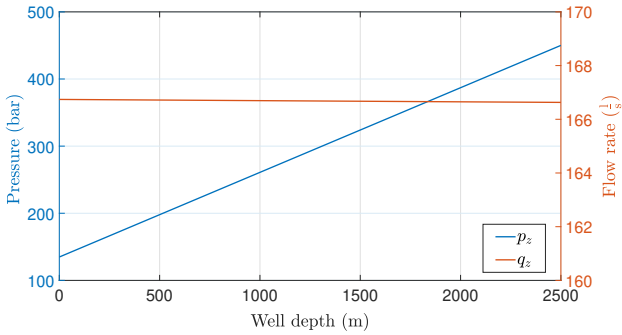
Figure 3. The pressure p_z and flow rate q_z along the well depth.

Figure 4. The model results, (a) reservoir pressure p_r and well pressure p_0, (b) flow through bit q_{bit}, bottom hole flow rate q_0 and return flow rate q_l, (c) control signal p_l, and (d) mud pit level h_m.

2.2.3 The Choke Controller

Based on the return flow estimates, the well flow and pressure estimates, and the estimated reservoir properties, a *desired* topside choke pressure is computed by the kick-/loss attenuation system. The kick/loss attenuation system is derived using the infinite-dimensional backstepping approach. In (Holta et al., 2017), stability is proved for the closed loop system consisting of the bottomside estima-tor and the pressure controller. From the choke equation (Equation 3), the variable choke opening $k_{choke}(t)$ can be adjusted to match a given return flow estimate and desired topside pressure.

3 Results and Discussion

The bottomside model and estimator are implemented in MATLAB using a 1st order upwind scheme for the spatial discretization and Euler scheme for the temporal discretization. The topside model and estimators are implemented in MATLAB using orthogonal collocation for the spatial discretization and Runge-Kutta 4th order scheme for the temporal discretization.

The simulation results of the well model, mainly the pressure p_z and flow rate q_z for constant input values can be seen from Figure 3. The linear relationship of pressure with the well depth can be clearly seen.

The simulation results of the combined model with a

Figure 5. The estimated reservoir pressure ($\hat{p}_r$) and BHP ($\hat{p}_0$) with the actual reservoir pressure (p_r).

Figure 6. The estimation error of BHP ($p_0 - \hat{p}_0$) and reservoir pressure ($p_r - \hat{p}_r$).
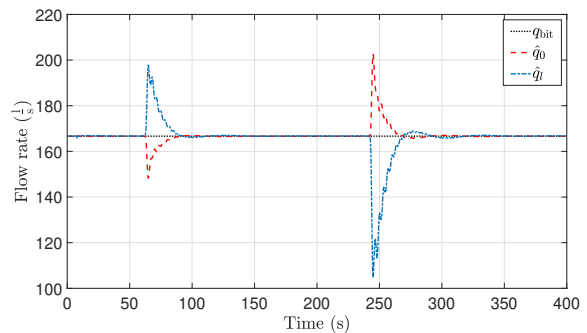
Figure 7. The estimated volumetric flow rates $\hat{q}_0$ and $\hat{q}_l$ against the flow through drill bit q_{bit}.

known reservoir pressure, and known flow through bit are shown in Figure 4. The objective of keeping the BHP close to the reservoir pressure is achieved with a small time lag of about 40 s as shown in (a) in Figure 4. Further, the control signal and the mud pit volume are shown.

The estimated BHP and the reservoir pressure using the bottomside adaptive observer are shown in Figure 5 together with the actual reservoir pressure.

The estimation error between the estimated BHP and the actual BHP is shown in Figure 6 as well as the estimation error between the estimated and actual reservoir pressure. The estimation error of the reservoir pressure is quite high due to the time lag. However, the BHP estimation error is comparatively low.
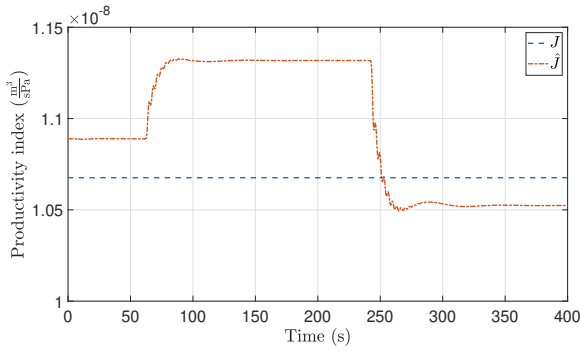
Figure 8. The estimated productivity index $\hat{J}$ and the constant productivity index J that is used in the model.

Figure 9. The estimated volumetric flow rates along the Venturi channel ($\hat{q}_{l_1}$ and $\hat{q}_{l_3}$) with the actual flow rate q_l.

Figure 10. The estimated active mud pit level $\hat{h}_m$ with the actual mud pit level h_m.
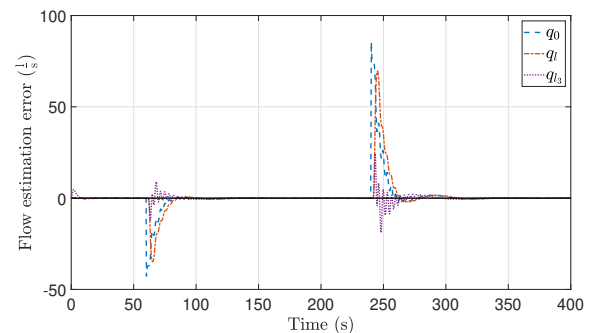
Figure 11. The estimation error of flow rate at bottom of the well ($q_0 - \hat{q}_0$), flow rate at top of the well ($q_l - \hat{q}_l$) and flow rate at the Venturi channel ($q_l - \hat{q}_{l_3}$), respectively.

Similarly, the observer results of the estimated volumetric flow rates at bottom and top of the well are shown in Figure 7 together with the volumetric flow rate through the drill bit. The change of the estimated parameter, productivity index, with the change of reservoir pressure can be seen in Figure 8 against the constant value used in the model.

The topside estimator results of the volumetric flow rates along the Venturi channel are shown in Figure 9 with the model result from the bottomside model. The numerical oscillations on the actual flow rate are filtered out by the UKF along the channel length as seen on the flow rate estimation of the third position of the channel. This flow rate is then used in the active mud pit model to estimate the mud pit level. The estimated mud pit level is shown in Figure 10 with the actual mud pit level. Since the flow rate fluctuation during disturbances are quite low, the mud pit level fluctuation is also comparatively low.

The estimation errors of different flow rate estimates are shown in Figure 11. The estimation error of the flow rate at the bottom of the well and the estimation error of flow rate at the top of the well are quite high similar to the pressure estimation error of the observer. However, the estimation error of the Venturi flow rate is comparatively low. Although the high estimation errors exist when considered separately, the estimated flow rate at the top of the well (return flow rate) for the complete system is taken to be the estimated Venturi channel flow rate.

4 Conclusions

The complete flow loop of a managed pressure drilling system is tested for a novel, mathematical model-based kick/loss detection scheme. A topside return flow estimator is combined with a bottomside well pressure and flow estimator to provide this complete kick/loss detection scheme. The topside flow rate estimator consists of a Venturi flowmeter, return flowline and the active mud pit. This return flow estimate is then used as input to the bottomside wellbore estimator. The distributed pressure and flow rate in the annulus, and the unknown reservoir pressure and the productivity index are estimated by an adaptive observer. The topside and bottomside estimators are then used in closed loop with a topside choke controller. The bottomhole pressure is controlled to be equal to the estimated reservoir pressure and the kick/loss is attenuated in a simulation environment. The estimation results are accurate and show improved kick/loss detection and attenuation capability.

Acknowledgement

A Appendix

The details of the estimators are briefly stated in this section.

A.1 Adaptive Observer

The system described in equations 1 - 5 can be written in an equivalent characteristic form using its Riemann coordinates (u, v), which are obtained through a linear transformation of variables $(p, q) \rightarrow (u, v)$ (see for example (Aamo, 2013)):

$$\frac{\partial u(z_n, t)}{\partial t} + \lambda \frac{\partial u(z_n, t)}{\partial z_n} = c_1(z_n) v(z_n, t) \tag{9}$$

$$\frac{\partial v(z_n, t)}{\partial t} - \mu \frac{\partial v(z_n, t)}{\partial z_n} = c_2(z_n) u(z_n, t) \tag{10}$$

$$u(0, t) = \theta_1 v(0, t) + \theta_2 \tag{11}$$

$$v(1, t) = U(t) \tag{12}$$

$$y(t) = u(1, t) \tag{13}$$

Here, $z_n \in [0, 1]$ and $t \geq 0$ are the independent variables, while $u(z_n, t)$ and $v(z_n, t)$ are the transformed states of the system. Further, $\lambda, \mu > 0$, and $c_1(z_n), c_2(z_n) \in C([0, 1])$ are known, while θ_1, θ_2 are unknown parameters given uniquely by the unknown reservoir parameters J, p_r. $U(t)$ and $y(t)$ are uniquely given by the topside pressure $p_l(t)$ and topside flow $q_l(t)$.

The following observer is designed for the system.

$$\frac{\partial \hat{u}}{\partial t} + \lambda \frac{\partial \hat{u}}{\partial z_n} = c_1(z_n) \hat{v} + P_1(z_n, t) \left[y(t) - \hat{u}(1, t) \right] \tag{14}$$

$$\frac{\partial \hat{v}}{\partial t} - \mu \frac{\partial \hat{v}}{\partial z_n} = c_2(z_n) \hat{u} + P_2(z_n, t) \left[y(t) - \hat{u}(1, t) \right] \tag{15}$$

$$\hat{u}(0, t) = \hat{\theta}_1 \hat{v}(0, t) + \hat{\theta}_2 \tag{16}$$

$$\hat{v}(1, t) = U(t) \tag{17}$$

Here, $\hat{u}, \hat{v}$ are the system state estimates, $\hat{\theta}_1, \hat{\theta}_2$ are the parameter estimates and P_1, P_2 are the output injection gains satisfying

$$P_1(x, t) = \lambda P^u(x, 1, t) \tag{18}$$

$$P_2(x, t) = \lambda P^v(x, 1, t) \tag{19}$$

where the kernels P^u, P^v are the unique solution to the system

$$P_t^u(x, \xi, t) + \lambda P_x^u(x, \xi) + \lambda P_\xi^u(x, \xi) = c_1(x) P^v(x, \xi) \tag{20}$$

$$P_t^v(x, \xi, t) - \mu P_x^v(x, \xi) + \lambda P_\xi^v(x, \xi) = c_2(x) P^u(x, \xi) \tag{21}$$

$$P^v(x, x) = \frac{c_2(x)}{\lambda + \mu} \tag{22}$$

$$P^u(0, \xi) = \hat{\theta}_1 P^v(0, \xi). \tag{23}$$

The equations are solved by a finite-difference method over a selected triangular discretized mesh of space.

The parameter estimates $\hat{\theta}_1, \hat{\theta}_2$ are generated by the adaptive laws

$$\dot{\hat{\theta}}_1(t) = \begin{cases} \Gamma \frac{\vartheta(t) - \bar{v}(t) \hat{\theta}_1 - \hat{\theta}_2}{2 + \bar{v}(t)} \bar{v}(t) & \text{for } t > t_F \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

$$\dot{\hat{\theta}}_2(t) = \begin{cases} \Gamma \frac{\vartheta(t) - \bar{v}(t) \hat{\theta}_1 - \hat{\theta}_2}{2 + \bar{v}(t)} & \text{for } t > t_F \\ 0 & \text{otherwise} \end{cases} \tag{25}$$

where

$$\vartheta(t) = y(t) - \hat{u}(1, t) + \hat{\theta}_1(t - d_\alpha) \bar{v}(t) + \hat{\theta}_2(t - d_\alpha) \tag{26}$$

$$\begin{aligned} \bar{v}(t) = &\hat{v}(0, t - d_\alpha) \\ &+ \int_0^1 P^v(0, \xi, -t - d_\alpha) y(t - \frac{\xi}{\lambda}) d\xi \\ &- \int_0^1 P^v(0, \xi, -t - d_\alpha) \hat{u}(1, t - \frac{\xi}{\lambda}) d\xi \end{aligned} \tag{27}$$

A.2 UKF

The shallow water equations (equations 6 and 7) can be simplified into ordinary differential equations using the orthogonal collocation (for example using 3 collocation points) as follows.

$$\frac{dA_{c_i}}{dt} = -\sum_{i=1}^{3} M_{ij}^T q_{l_i} \tag{28}$$

$$\begin{aligned} \frac{dq_{l_i}}{dt} = &-\sum_{i=1}^{3} M_{ij}^T \frac{q_{l_i}^2}{A_{c_i}} - g \sum_{i=1}^{3} M_{ij}^T I_{1_i} \cos \phi \\ &+ g A_{c_i} \sin \phi - T_{f_i} \end{aligned} \tag{29}$$

$$M = \frac{1}{l_c} \begin{bmatrix} -3 & 4 & -1 \\ -1 & 0 & 1 \\ 1 & -4 & 3 \end{bmatrix}, \tag{30}$$

Here $i, j \in [1, 2, 3]$. The detailed derivation of the equations can be found in (Jinasena et al., 2017). This sim-plified non-linear system for the UKF can be written in discrete time form as follows,

$$X_{k+1} = f_n(X_k, t_k) + w_k(t_k), \tag{31}$$

$$y_k = CX_k + v_k(t_k), \tag{32}$$

The function f_n represents the nonlinear model given by Equations (28)–(30) and (8) for $\dot{h}_i$, $\dot{q}_{l_i}$, and $\dot{h}_m$, respectively. k is the discrete time index. The measurement noise in the measured output at t_k is denoted by $v_k \in \mathbb{R}^{n_m}$, where $v_k \sim \mathcal{N}(0, R)$. Similarly, $w_k \in \mathbb{R}^{n_s}$, where $w_k \sim \mathcal{N}(0, Q)$ accounts for process noise. Here, n_m and n_s are the number of measurements and number of states, while R and Q are the covariance matrices for measurement noise and process noise, respectively.

The states and measurements can be chosen from the available fluid levels as preferred, based on the

necessity and observability. For example; $X = [h_1\ h_2\ h_3\ q_{l_1}\ q_{l_2}\ q_{l_3}\ h_m]^T$ and $y = [h_2\ h_3\ h_m]^T$ is chosen for faster convergence of the observer, thus C becomes as follows,

$$C = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \quad (33)$$

The estimated return flow rate $\hat{q}_{l_1}$ or $\hat{q}_{l_3}$ is then used in the bottomside estimator.

Nomenclature

α	Momentum correction coefficient
β	Bulk modulus of the drilling mud
λ, μ	Eigen values
ϕ	Bottom slope angle of the channel
ρ	Density of the drilling mud
θ_1, θ_2	Boundary parameters
A_a	Cross sectional area of the annulus
A_c	Wetted cross sectional area of the channel
A_m	Cross sectional area of the active mud pit
C	Measurement matrix of the UKF
c_1, c_2	Design parameters
F_a	Friction factor of the annulus
f_n	A nonlinear function
g	Acceleration of gravity
h	Fluid level
h_m	Active mud pit level
I_1	First moment of area
I_2	Pressure forces in the fluid volume
J	Productivity index
k	Discrete time index
k_{choke}	Choke coefficient
l	Length
l_c	Length of the channel
l_w	Well depth

M	A matrix
n_m	Number of measurements, UKF
n_s	Number of states, UKF
p	Pressure
P_1, P_2	Output injection gains
p_0	Atmospheric pressure
p_r	Reservoir pressure
Q	Covariance matrix for process noise
q	Volumetric flow rate
q_{bit}	Volumetric flow through the drill bit
q_{loss}	Fluid losses at solid removal equipment
q_{pump}	Mud pump flow rate
R	Covariance matrix for measurement noise
t	Time
T_f	Non-Newtonian friction term
U	Control input
u	Transformed state, pressure
v	Transformed state, flow rate
v_k	Measurement noise
w_k	Process noise
X	States of the UKF
x	Position along the channel length
y	Measurement vector
z	Position along well depth
z_n	Normalized position along well depth

References

Ole Morten Aamo. Disturbance Rejection in 2×2 Linear Hyperbolic Systems. *IEEE Transactions on Automatic Control*, 58(5):1095–1106, 2013. doi:10.1109/TAC.2012.2228035.

Christian Berg. *Modeling for Automatic Control and Estimation of Influx and Loss during Drilling Operations*. PhD, University of South-Eastern Norway, 2020.

Christian Berg, Anjana Malagalage, Cornelius E Agu, Glenn-Ole Kaasa, Knut Vaagsaether, and Bernt Lie. Model-based Drilling Fluid FLow Rate Estimation Using Venturi Flume. *IFAC-PapersOnLine*, 48(6):171–176, 2015. doi:10.1016/j.ifacol.2015.08.027.

M. Hanif Chaudhry. *Open-Channel Flow*. Springer, New York, 2nd edition, 2008. ISBN 9780387301747.

Khim Chhantyal, Hakon Viumdal, Saba Mylvaganam, and Geir Elseth. Estimating viscosity of non-Newtonian fluids using support vector regression method: Rheological parameters of drilling fluids using data fusion. *SAS 2016 - Sensors Applications Symposium, Proceedings*, pages 285–290, 2016. doi:10.1109/SAS.2016.7479860.

Khim Chhantyal, Morten Hansen Jondahl, and Håkon Viumdal. Upstream Ultrasonic Level Based Soft Sensing of Volumetric Flow of non-Newtonian Fluids in Open Venturi Channels. *IEEE Sensors Journal*, 18(12):5002–5013, 2018. doi:10.1109/JSEN.2018.2831445.

Ven Te Chow. *Open-Channel Hydraulics*. McGraw-Hill, New York, 1959. ISBN 0070107769.

L. P. Dake. *Fundamentals of Reservoir Engineering*. Elsevier, Amsterdam, 17th edition, 1998. ISBN 0-444-41830-X.

Mohamed S. Ghidaoui, Ming Zhao, Duncan A. McInnis, and David H. Axworthy. A review of water hammer theory and practice. *Applied Mechanics Reviews*, 58(1):49–76, jan 2005. doi:10.1115/1.1828050.

Haavard Holta, Henrik Anfinsen, and Ole Morten Aamo. Adaptive Set-Point Regulation of Linear 2x2 Hyperbolic Systems with Uncertain Affine Boundary Condition using Collocated Sensing and Control. In *2017 Asian Control Conference, (ASCC)*, pages 2766–2771, Australia, 2017. doi:10.1109/ASCC.2017.8287615.

Asanthi Jinasena. *Models and Estimators for Flow of Topside Drilling Fluid*. PhD Thesis, University of South-Eastern Norway, 2019.

Asanthi Jinasena and Roshan Sharma. Estimation of Mud Losses during the Removal of Drill Cuttings in Oil Drilling. *SPE Journal*, 2020. doi:10.2118/201230-PA. (in press).

Asanthi Jinasena, Glenn-Ole Kaasa, and Roshan Sharma. Use of Orthogonal Collocation Method for a Dynamic Model of the Flow in a Prismatic Open Channel : For Estimation Purposes. In *Proceedings of the 58th Conference on Simulation and Modelling (SIMS 58)*, pages 90–96, Reykjavik, Iceland, 2017. Linköping University Electronic Press. doi:10.3384/ecp1713890.

Asanthi Jinasena, Ali Ghaderi, and Roshan Sharma. Modeling and Analysis of Fluid Flow through A Non-Prismatic Open Channel with Application to Drilling. *Modeling, Identification and Control*, 39(4):261–272, 2018. doi:10.4173/mic.2018.4.3.

Asanthi Jinasena, Glenn-Ole Kaasa, and Roshan Sharma. Improved Real–Time Estimation of Return Flow Rate of Drilling Fluids by Model Adaptation for Friction Parameter. *IEEE Sensors Journal*, 19(20):9314–9323, 2019. doi:10.1109/jsen.2019.2923854.

Morten Hansen Jondahl. *Data Driven Models for Estimation of Drilling Fluid Rheological Properties and Flow Rate*. PhD Thesis, University of South-Eastern Norway, 2020.

Morten Hansen Jondahl and Håkon Viumdal. Estimating Rheological Properties of Non-Newtonian Drilling Fluids Using Ultrasonic-Through-Transmission Combined with Machine Learning Methods. *IEEE International Ultrasonics Symposium, IUS*, October(255348):1–4, 2018. doi:10.1109/ULTSYM.2018.8579796.

Morten Hansen Jondahl and Håkon Viumdal. Developing Ultrasonic Soft Sensors to Measure Rheological Properties of Non-Newtonian Drilling Fluids. *Technisches Messen*, pages 1–14, 2019. doi:10.1515/teme-2019-0039.

Ivan Pirir, Asanthi Jinasena, and Roshan Sharma. A Dynamic Model for Drain Back to Active Mud Pit Combined with a Well Model during Drilling. *Journal of Petroleum Science and Engineering*, 167:803–818, 2018. doi:10.1016/j.petrol.2018.04.057.

Prasanna Welahettige. *Transient drilling fluid flow in Venturi channels : comparing 3D and 1D models to experimental data*. PhD Thesis, University of South-Eastern Norway, 2019.

Prasanna Welahettige, Christian Berg, Joachim Lundberg, Bernt Lie, and Knut Vaagsaether. Computational Fluid Dynamics Study of the Effects of Drill Cuttings on the Open Channel Flow. *International journal of Chemical Engineering*, 2019: 1–9, 2019.

Adaptive Response Time for Integrated Autodriller Controller

Khim Chhantyal   Kjartan Thor Birgisson   Akash Padir*

National Oilwell Varco

khim.chhantyal@nov.com kjartan.birgisson@nov.com
*Correspondence: akash.padir@nov.com

Abstract

The integrated autodriller controller (IADC) is a software product by National Oilwell Varco (NOV) developed to allow an operator to drill with a constant weight on bit (WOB). The control loop has a proportional-integral (PI) controller, and PI gains are dependent on drill string compliance and response time (RT). The compliance is calculated based on the drill string length, but RT is difficult to predict. The existing IADC uses an empirical regression model to estimate RT based on the string length, which is not optimal during a formation change. This work presents an adaptive response time (ART), a machine learning model to predict RT based on the rate of penetration (ROP). ART can adapt to formation changes and gives smoother regulation of block velocity and WOB. Simulation results show a better torque and speed regulation of drill string using ART.

Keywords: oil & gas drilling, integrated autodriller controller, adaptive response time, machine learning

Abbreviations: ART - Adaptive Response Time, CM - Comprehensive Model, IADC - Integrated Autodriller Controller, NOVOS - National Oilwell Varco Operating System, PI - Proportional-Integral, ROP - Rate of Penetration, RPM - Revolutions per Minute, RT - Response Time, TD - Top Drive WOB - Weight on Bit

1 Introduction

The ability to maintain constant weight on bit (WOB), torque, and revolutions per minute (RPM) is essential to achieve optimal drilling efficiency for conventional drilling rigs. Control over these parameters maximizes bit life and increases ROP, thus resulting in reduced drilling completion time. Autodrillers are the commonly used tools in the drilling industry for WOB control by applying regulated force on the drill bit by means of a proportional-integral (PI) controller. An industry-wide problem with autodrillers is maintaining regulation when drilling from one formation to another of different strength (Pastusek et al., 2016; Badgwell et al., 2018; Cayeux, 2018). This causes the error within the control loop to suddenly increase or decrease, resulting in controller instability. (Pastusek et al., 2016) pointed out the coupling between incorrect autodriller regulation and the drill string stick-slip. Operator intervention is required to stabilize the regulation, either by changing the setpoint or gain of the con-

troller (Pastusek et al., 2016; Badgwell et al., 2018; Adam, 2018). Therefore making the success of the process dependent on the operator's skills.

In recent years, there have been several studies done to tackle this problem as the industry is shifting with an increased focus on autonomous drilling and limiting the need for human intervention. Adam (2018) collected autodriller parameters and corresponding ROP from successful drilling operations and created a road map intended to assist operators to select optimal autodriller parameters. Badgwell et al. (2018) proposed to use the ROP setpoint to auto-adjust the WOB gain to respond to autodriller instabilities. In addition, Badgwell et al. (2018) challenged autodriller vendors to address the WOB controller issues. We accepted this challenge.

Well dynamics are complex and well characteristics differ drastically between geological areas. Furthermore, lack of downhole sensors forces controllers to be regulated indirectly from top side equipment. Thus, developing general solutions with principle mechanistic models can be complicated. We therefore decided to use a data-driven approach like (Adam, 2018) in combination with the existing mechanistic model. Data-driven methods require data, which in the drilling industry is often proprietary. Obtaining a diverse dataset to ensure correct response from a machine learning algorithm in every scenario is also not feasible. For this reason, we used the approach of generating data by means of simulation.

2 Methods

2.1 Simulation Environment Set-up

The simulation environment consists of NOV's simulated integrated autodriller controller (IADC) and NOV's comprehensive model (CM). The IADC is an autodriller designed for regulating WOB, ROP, differential pressure, and torque limit. For this simulation, only the WOB mode (softWOB^TM) of the IADC is used. The CM is a simulation tool for drilling operations that contains fluid dynamics, string dynamics, formation characteristics, top side equipment, etc. to simulate well bore behavior. The IADC receives the WOB, ROP, top drive torque, and block position from the CM. Depending on setpoints and limits, the IADC adjusts the block velocity and feeds to the CM. Figure 1 shows a simplified block diagram of the simulation environment.
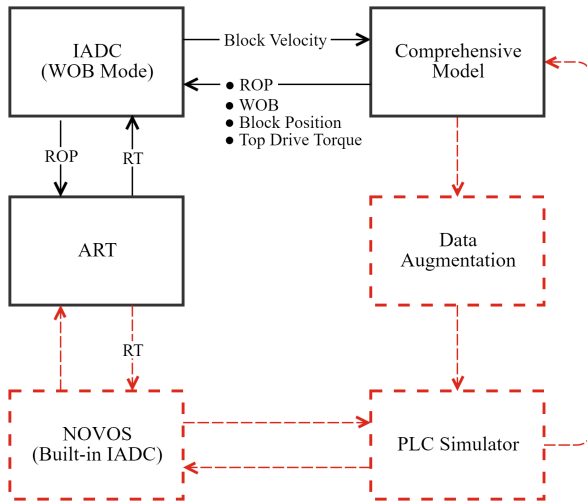
Figure 1. A simplified block diagram of the simulation environment. Current environment is displayed in black color, the NOVOS-based environment is displayed in red color (dotted lines).

Using the simulation environment, we created a dataset by performing drilling simulations through different formations in combination with various gains and depths. The dataset is used to train a machine learning based, Adaptive Response Time (ART) model. ART receives ROP as the only input from the IADC and predicts the response time (RT) for the controller. The IADC is currently executed in a PLC environment. The proposed ART is a polynomial model that can be implemented with minimal changes in existing installations and is ready for field testing.

ART is likely to improve the current IADC in a wide range of drilling scenarios. However, the problem is multi-dimensional and non-linear; therefore, the single input polynomial model is limited. To further improve ART, the feature space needs to be increased and a non-linear model implemented. This requires more drilling parameters and higher computational power. NOVOS is NOV's high level operating system, an automation platform built for full automatic control of the drilling process. Within NOVOS, ART has access to all possible drilling parameters and the computational power needed for training and executing non-linear models. The current simulation environment comprises of multiple mechanistic models. To increase the realism of the simulation model, we have added a simulator for all the equipment PLCs and a data augmentation layer for the signals coming from the CM to augment signal noise and variations. This environment will be used for simulations with the ART implemented in NOVOS. The NOVOS-based simulation environment is highlighted with red color (dotted lines) in Figure 1.

2.2 Auto-Tune IADC

IADC uses a PI controller to regulate WOB. Equation 1 shows the standard form of a PI controller.

$$u(t) = K_p e(t) + \frac{K_p}{T_i} \int_0^t e(t') dt' \qquad (1)$$

For IADC, the control variable (u) is the block velocity and error (e) is the difference between setpoint and calculated WOB. The proportional and integral gain parameters are function of drill string compliance and response time.

The drill string compliance is calculated based on the drill string length and mass. RT is a tunable time constant and can be set between 0.5 [s] and 20 [s]. It can be interpreted as a time the controller should use to adjust the block velocity to compensate for the error. As a general rule, RT should be tuned every 100-500 meters. Setting too short response time (closer to 0.5) implies an aggressive controller. This may cause the IADC to overreact on WOB changes and leads to increasingly large overshoots and oscillations. A high value of RT (closer to 20) implies slower response to WOB changes. It can tackle overshooting and reduce oscillations; however, it can take a long time to reach setpoint.

IADC currently offers functionality to estimate response time based on the drill string length. The function is a regression model based on empirical data from model analysis as well as real-world experiments. This approach does not account for formation type being drilled, resulting in unstable regulation caused by formation change. ART is designed to tackle this issue by detecting current formation using calculated ROP.

3 Results

In this section, simulation results of the IADC with traditional approach and with ART are presented. Table 1 shows the drilling parameters and ranges used in this study. As per traditional approach, RT (2.55 seconds) is selected based on the regression model for the given bit depth (2,050 meters).

Table 1. Drilling parameters and their ranges, used in the simulations.

Parameters	Range
Drilling depth	2,040 - 2,055 [m]
ROP limit	$\pm$ 170 [m/h]
WOB limit	$\pm$ 25 [tons]
WOB setpoint	20 [tons]
Drillability	15 - 66 [m/h]

To analyse the simulation results, the following parameters are plotted against time in the subsequent subsections.

- WOB: operator setpoint vs. measured weight on bit

- MeasuredDepth: drilling progress against time

- Drillability: parameter describing relation between rock toughness and the bit's ability to produce hole.
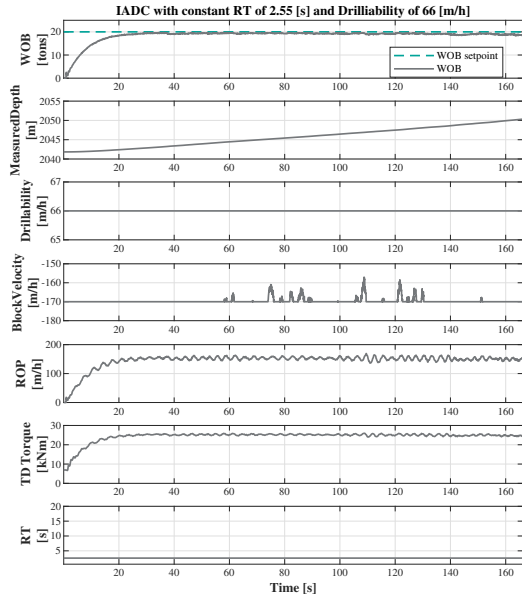
Figure 2. The performance of the IADC with constant RT = 2.55 [s] and constant drillability = 66 [m/h].

A higher number results in faster drilling (higher ROP) for a given weight/speed on bit.

- BlockVelocity: block velocity due to controller action (positive/negative represents hoisting/lowering)

- ROP: calculated rate of penetration

- TDTorque: measured top drive torque

- RT: response time used by the controller

3.1 Case 1: IADC performance in homogeneous formation

In this case, a homogeneous formation is chosen with drillability of 66 [m/h] (indicating soft formation) throughout the drilling simulation. The value of drillability is chosen such that it suits the traditionally calculated RT (= 2.55 [s]). Figure 2 shows that the WOB smoothly follows setpoint. In this ideal drilling set-up, acceptable torque and speed regulation are achieved.

3.2 Case 2: IADC performance against changing formations

For this simulation, the drilling set-up is maintained as in Case 1 with the exception of formation. Here, formation is simulated to change from soft (66 [m/h]) to harder formation (50 [m/h]). It is noticed that the constant RT = 2.55 [s] does not suit the harder formation. As the RT is not suitable, it results in block velocity oscillations, causing oscillations in WOB, ROP, and top drive torque as shown in Figure 3. The results indicates that the controller needs to be tuned in an event of formation change.

By adjusting controller parameters (i.e., WOB setpoint or RT), the operator can overcome oscillations caused by

Figure 3. The performance of the IADC with constant RT = 2.55 [s] and changing drillability from 66 to 50 [m/h].

formation change. In Figure 4, WOB setpoint is reduced from 20 [tons] to 10 [tons] and to 5 [tons] by the operator after noticing the oscillations. The oscillations reduce with changes in WOB setpoint. As the ROP is directly dependent on WOB, this results in reduced ROP. Hence, a proper way to reduce or eliminate oscillations and maintaining the ROP is by tuning RT for changing formation.
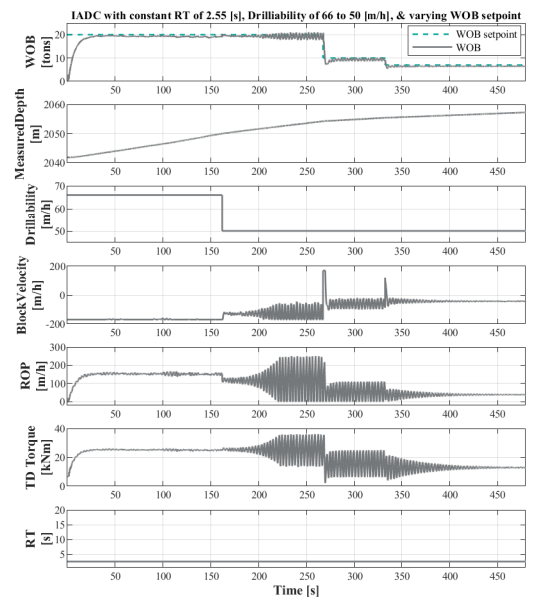
Figure 4. The performance of the IADC with constant RT = 2.55 [s] and changing drillability from 66 to 50 [m/h]. WOB setpoint is reduced to overcome the oscillations, which results in reduced ROP.
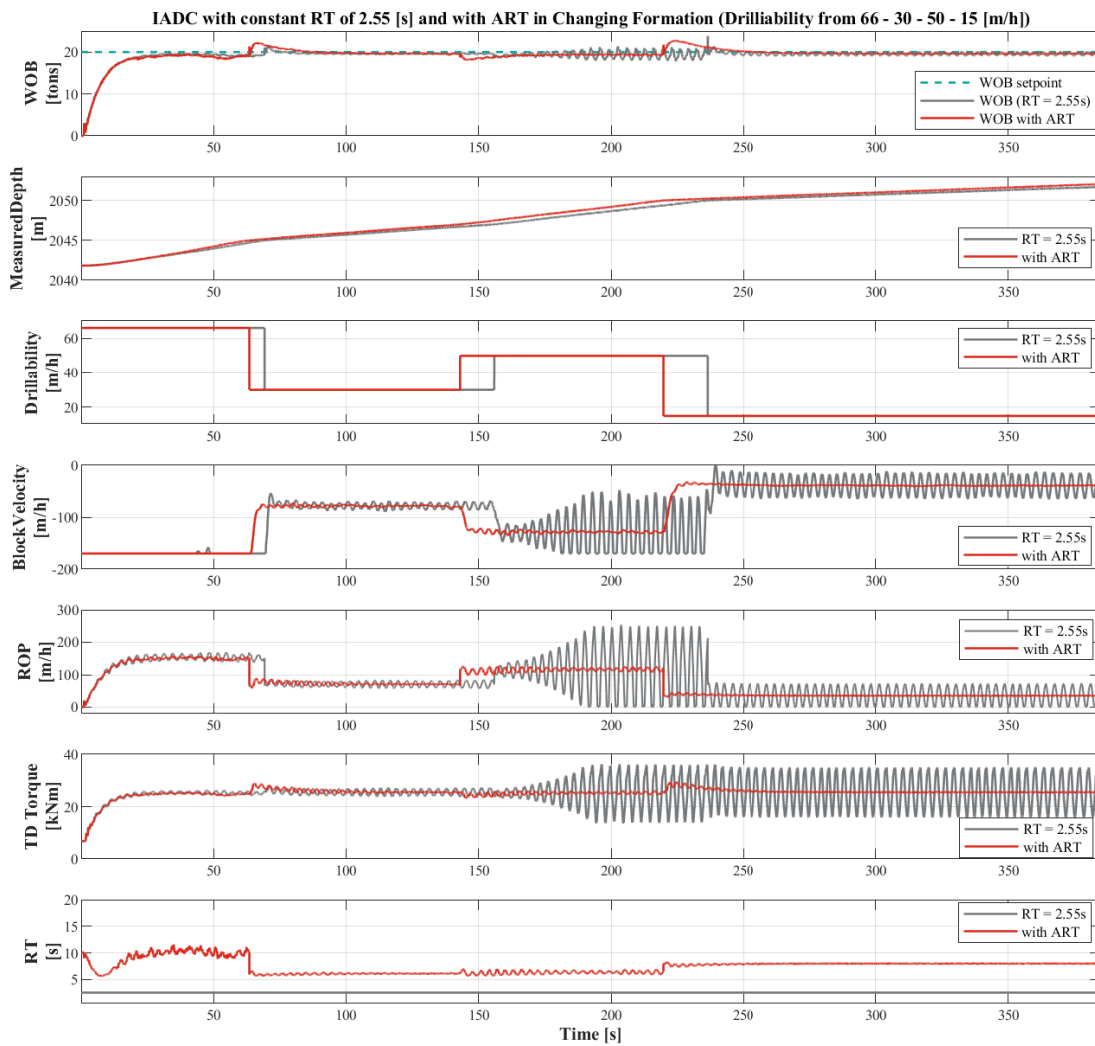
Figure 5. A performance comparison of the IADC with constant RT and with ART. IADC with ART can adapt to formation changes and provides smoother WOB, block velocity, and top drive torque. Stable top drive torque indicates elimination of autodriller induced stick-slip.

3.3 Case 3: IADC with ART

In Figure 5, drillability is randomly varied from 66 to 15 [m/h] with increasing depth. ART can adapt to any formation change and selects a suitable RT based on ROP. The IADC with ART provides better regulation of WOB and stable block velocity. Improved regulation also causes top drive torque to stabilize, which is an indication of reduced stick and slip of the drill string. This ultimately results in increased drilling efficiency (increased ROP).

4 Conclusion

In the event of formation change, the existing regulator undergoes oscillations. These oscillations can be overcome by changing controller parameters (i.e., setpoint and gains). The current practice of manually changing the setpoints to tackle the oscillations, can lead to reduced drilling efficiency. Therefore, the changing formation should be addressed by tuning the response time. The presented machine learning algorithm based ART is capable of detecting the change in formation (by utilizing calculated ROP) and is able to estimate a suitable RT. During simulations, ART is found to stabilize the process to great extent, resulting in improved drilling efficiency.

Currently, ART is trained with the limited drilling range (1,000 - 3,000 meters depth). The presented simulations are performed within the trained range, but with different simulation set-up. ART further needs to be trained for the entire possible drilling range (up to 10,000 meters depth). In addition, ART can further be improved by adding features to the input feature space of the machine learning algorithm and with implementation of non-linear models in a high-level programming environment.

5 Acknowledgment

The authors would like to thank NOV for the permission to publish this paper. As well as NOV's Robotic Control team, NOV's NOVOS team, & NOV's R&D for the collaboration.

Special thanks to Matthew Jackson, Francisco Ocegueda-Hernandez, Ivan Rajic, Njål Sama, Charles Holst Sandnes, and Morten Ørevik for their support.

References

D Adam. Reducing stick-slip by avoiding auto-driller control dysfunction. In *IADC/SPE Drilling Conference and Exhibition*. Society of Petroleum Engineers, 2018. doi:10.2118/189653-MS.

T Badgwell, P Pastusek, and K Kumaran. Auto-driller automatic tuning. In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers, 2018. doi:10.2118/191417-MS.

E Cayeux. On the importance of boundary conditions for real-time transient drill-string mechanical estimations. In *IADC/SPE Drilling Conference and Exhibition*. Society of Petroleum Engineers, 2018. doi:10.2118/189642-MS.

P Pastusek, M Owens, D Barrette, V Wilkins, A Bolzan, J Ryan, K Akyabi, M Reichle, and D Pais. Drill rig control systems: Debugging, tuning, and long term needs. In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers, 2016. doi:10.2118/181415-MS.

Near-well simulation of oil production from a horizontal well with ICD and AICD completions in the Johan Sverdrup field using OLGA/ROCX

Ali Moradi, Britt Margrethe Emilie Moldestad

Department of Process, Energy and Environmental Technology, University of South-Eastern Norway, Norway.
{ali.moradi@usn.no, britt.moldestad@usn.no}

Abstract

One of the main principles of improving oil recovery is maximizing the contact area between the well and the reservoir. To achieve this purpose especially in reservoirs with a thin oil column, long horizontal wells are widely used today. However, there are some challenges related to horizontal wells like water coning towards the heel due to the heel-toe effect as well as early water breakthrough owing to heterogeneity along the well. In order to tackle these issues, passive inflow control devices (ICDs) and autonomous inflow control devices (AICDs) can be used. ICDs are able to balance the drawdown pressure along the horizontal well and as a result, postpone the early water breakthrough. By applying AICDs, in addition to postponing the early water breakthrough, water can be partially choked back autonomously, and the negative impacts of early water breakthrough will be attenuated. The Johan Sverdrup field (JSF) is a giant oil field located in the North Sea and production from this field has been started recently. Since there is a plan for developing this oil field in the near future, and a few studies have been done on this field so far, further studies are needed to obtain more cost-effective oil recovery in this field. The main objective of this paper is near-well simulation of oil production from the well 16/2-D-12 in the JSF by considering ICD and AICD completions. The simulation has been conducted based on the characteristics of the reservoir near this well for 750 days of oil production. OLGA in combination with ROCX has been used as a simulation tool. The simulation results showed that by applying both ICDs and AICDs the heel-toe effect, and heterogeneity along the well can be effectively handled and the water breakthrough time can be delayed by 255 days. Moreover, it was observed that by completion of the well 16/2-D-12 with AICDs, the accumulated water production can be reduced by 11.9% compared to using ICDs. In the same way, by using AICDs the flow rate of water production is reduced by 13.4% after 750 days. Furthermore, the results showed that using AICDs has a negligible impact on both the accumulated and the flow rate of oil production compared to using ICDs. Therefore, by completion of the well 16/2-D-12 with AICDs more cost-effective oil production can be achieved.

1 Introduction

Despite the rapid progress in the area of renewable energy, fossil fuels will remain the world's most important energy source for the years to come. Therefore, in order to meet future energy demands, increasing oil production and reducing carbon footprint from the oil and gas sector must be in the spotlight. Increasing oil production requires massive investments to explore and develop new oil fields. For stimulating new investments in the oil industry, new technologies and methods must be applied for increasing the efficiency of oil production and achieving maximum profit per barrel of oil. Oil production from the Johan Sverdrup field has been started recently. With the aim of obtaining maximum oil recovery, cutting-edge technologies have been applied in this field and further studies are needed to improve these technologies. The well 16/2-D-12 is a horizontal well located in the JSF and compared to other wells in this field, there is more information available related to the characteristics of the reservoir near this well. Therefore, this paper focuses on making a model for near-well simulation of oil production from the well 16/2-D-12. By using this simulation model, the functionality of ICDs and AICDs in improving oil recovery from the JSF is investigated.

Preparing a realistic model for the simulation of oil production needs to know the realistic characteristics of the reservoir. Since the extraction of oil from the JSF has just started, there is limited information available in the literature describing the characteristics of this field. As a result, determining some characteristics of the field like the viscosity of oil in the reservoir, anisotropy of permeability, wettability and relative permeability, etc. is not straightforward and needs curve-fitting on existing data, calculation by using general equations or some close to the mark assumptions. Moreover, due to the lack of information about the heterogeneity of the reservoir along the well with respect to permeability, the model is prepared for a heterogeneous reservoir with some random high permeable zones in different directions. In the following subchapters, the main steps for developing the simulation model as well as the simulation results are described and discussed.

2 Inflow control technologies

The horizontal well technology is wildly used for increasing the efficiency of oil recovery. One of the major problems that can reduce the efficiency of using horizontal wells is early gas and water coning. This problem gives rise to an uneven inflow profile along the horizontal well. Inflow control technologies are used to counteract the non-uniform inflow throughout the length of the horizontal well leading to improve the well performance and efficiency.

2.1 Passive Inflow Control Devices

One of the mature technologies that have been developed for managing the early water breakthrough in horizontal wells is passive Inflow Control Devices (ICDs). A passive ICD is a flow restrictor device with no moving part installed as a part of the sandface completion hardware. ICDs are used for chocking the flow by adding extra pressure drop and thereby even out the flow influx along a horizontal well. As Figure 1 is showing, ICDs can delay the early water breakthrough by balancing the inflow along the well and as a result, increasing the oil production (Aakre, 2017).
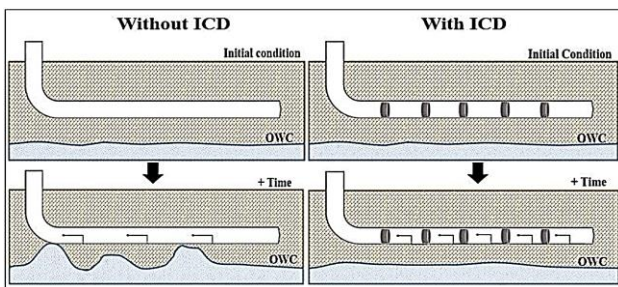
Figure 1. Application of ICDs in mitigation of the early water breakthrough (Chammout et al, 2017).

One of the main drawbacks of passive ICDs is that they are not able to choke the water back after breakthrough. In this condition, the well must be shut in to avoid increasing the water cut more than the capacity of the separation facilities and it leads to the reduction of oil production (Aakre, 2017).

Horizontal wells are completed with many ICDs distributed along the well. The pressure drop across an ICD is a function of flow rate, the geometry of ICD, and the fluid density but it does not depend on the fluid viscosity. The mathematical equation governing the behavior of orifice-type ICDs is:

$$\dot{Q} = C_D A \sqrt{\frac{1}{1 - \beta^4}} \cdot \sqrt{\frac{2\Delta P}{\rho}} \qquad (1)$$

where $\dot{Q}$ is the volume flow rate of the fluid passing through the ICD, ΔP is the pressure drop over the ICD, ρ is the fluid density and $\beta = d / D$ (d and D are the diameters of the orifice and production tubing respectively). In Equation 1, C_D is called *discharge*

coefficient. It can be calculated as $C_D = A / A_{vc}$ in which A is the cross-sectional area of the orifice hole and A_{vc} is the minimum jet area just downstream of the orifice called *Vena Contracta* (The Engineering ToolBox, 2004).

2.2 Autonomus Inflow Control Devices

Since passive ICDs have no ability to choke the water or gas back after breakthrough, Autonomous Inflow Control Devices (AICDs) have been developed as a robust alternative in recent years. AICDs have a moveable disk and they can be partially closed for low viscous fluids compared to oil like water and gas. As a result, in addition to delaying the water or gas breakthrough, AICDs can reduce the production of water or gas after breakthrough autonomously and increase oil production compared to passive ICDs with no need to control from the surface (Aakre, 2017).

AICDs are available with different designs and one of the most widely used types of AICDs has been developed by Norsk Hydro and Statoil and is called Rate Control Production (RCP). As Figure 2 shows, this type of AICD consists of three parts, which are the valve body, the nozzle, and the free-floating disk. The valve is designed based on the fluid properties in such a way that when oil passes through the valve, the pressure at the inlet is higher and the disk rests at the seat. Thus, the maximum flow area is obtained and as a result, the maximum amount of oil passes through the valve. Besides, the position of the moveable disk can vary based on the fluid properties and flow conditions. Owing to the special design of this valve, when low viscous fluids like water or gas enter the valve, the pressure at the inlet becomes lower. Consequently, based on Bernoulli's equation the total force acting on the disk pulls it towards the inlet, and the flow area is partially closed. Therefore, RCPs can minimize the flow rate of unwanted fluids like water or gas autonomously (Mathiesen et al, 2011; Askvik and Sørheim, 2017).
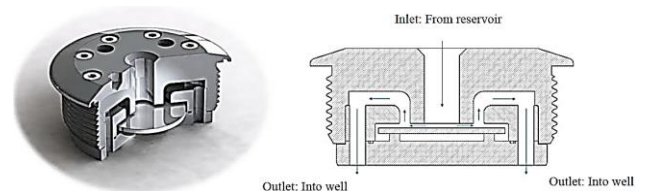
Figure 2. Statoil's RCP valve and its schematic sketch (Mathiesen et al, 2011).

The pressure drop across an RCP valve can be determined by an empirical mathematical function developed and validated by Statoil, and it is represented as:

$$\Delta P = f(\rho, \mu) \cdot a_{AICD} \cdot \dot{Q}^x \qquad (2)$$

where $\dot{Q}$ is the volumetric flow rate of fluid, and a_{AICD} and x are user-input parameters depending on the RCP

design and fluid properties. $f(\rho,\mu)$ is an analytical function of fluid density and viscosity and defined as:

$$f(\rho,\mu) = \left(\frac{\rho_{mix}^2}{\rho_{cal}}\right) \cdot \left(\frac{\mu_{cal}}{\mu_{mix}}\right)^y \qquad (3)$$

where y is a user-input parameter, and ρ_{cal} and μ_{cal} are specified as calibration density and viscosity respectively. Moreover, ρ_{mix} and μ_{mix} are the density and viscosity of the mixture of fluids passing through the RCP valve and are defined by:

$$\rho_{mix} = \alpha_{oil}\rho_{oil} + \alpha_{water}\rho_{water} + \alpha_{gas}\rho_{gas}$$
$$\mu_{mix} = \alpha_{oil}\mu_{oil} + \alpha_{water}\mu_{water} + \alpha_{gas}\mu_{gas} \qquad (4)$$

where α_{oil}, α_{water} and α_{gas} are the volume fraction of oil, water and gas in the mixture respectively (Halvorsen et al, 2016).

3 Characteristics of the reservoir near the well 16/2-D-12

To prepare a realistic model for simulation of oil production from the well 16/2-D-12 in OLGA/ROCX it is necessary to know the realistic characteristics of the reservoir near the drainage area of this well. Based on the well test data, the temperature of the reservoir near the well 16/2-D-12 is 81°C (354 K). Moreover, by using the reported formation pressure data, the formation pressure near this well can be approximated to be about 240 bar.

There is no information about the viscosity of crude oil at the reservoir conditions near the well 16/2-D-12. However, its value is known in temperatures 293 K, 313 K, and 323 K and they are given in the Equinor's crude summary report (Equinor, 2020). Based on the available values, the value of crude oil viscosity at the reservoir conditions (the pressure of 240 bar and the temperature of 354K) can be extrapolated. The method used for extrapolation is based on the linear regression and a commonly used mathematical model between temperature and viscosity described by Equation 5. Figure 3 illustrates the fitted curve and the viscosity of oil at the reservoir temperature calculated by MATLAB.

$$\mu = Ae^{B/T} \qquad (5)$$

By using curve-fitting on the known values of viscosity and the well test data provided by NPD, the given information in Table 1 for crude oil properties has been collected (Norwegian Petroleum Directorate, 2020).

Table 1. Crude oil properties near the well 16/2-D-12.

Parameter	Value
API gravity	28.0
Viscosity	3 cP
GOR	44 Sm^3/Sm^3
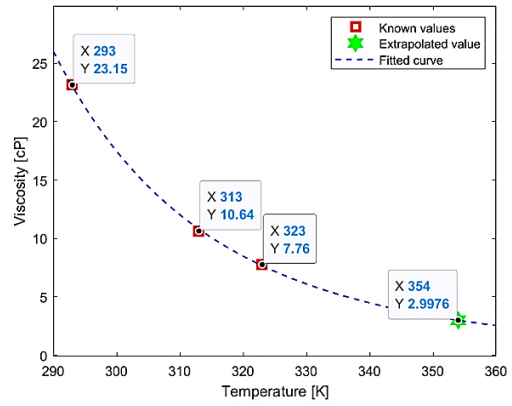
Figure 3. Estimation of crude oil viscosity at the reservoir temperature by curve-fitting on the known values.

According to the NPD's well test data obtained from the reservoir near the well 16/2-D-12 (Norwegian Petroleum Directorate, 2020), the main reservoir has an oil column of 67.5 m started from the depth of 1877 m and ended to the depth of 1944.5 m. Moreover, the reservoir is located in two zones with different types of sandstone. From the depth of 1877 m to 1911 m, the reservoir is located in the Intra Draupne sandstone formation, and between the depth of 1911 m and 1944.5 m, it is mostly located in the Vestland Group sandstones. Based on the obtained information from the analysis of the well log data near the well 16/2-D-12, the thickness of the net pay zone, the shale volume, and water saturation of the reservoir near this well have been calculated and summarized in Table 2 (Kaspersen, 2016).

Table 2. Thickness of the net pay zone, average water saturation, and shale volume near the well 16/2-D-12.

Zone	Net Pay Thick.	Average S_w	Average V_{sh}
Zone 1	32.46 m	0.062	0.09
Zone 2	22.86 m	0.322	0.122

Based on the analysis of well logs, the average effective porosity for both zones of the reservoir near the well 16/2-D-12 is 0.27. The median permeability near this well is 14.7 D but it varies between 0.5 and 40 D throughout the JSF. (Kaspersen, 2016).

By using the given data in Table 2, and Equations 6,7 and 8 which are empirical correlations for the sandstone reservoir, the anisotropy permeability, $a = k_v/k_H$, near the well 16/2-D-12 can be calculated (Iheanacho et el, 2012).

$$k_H = \sqrt{k_x k_y} \qquad (6)$$

$$k = \sqrt[3]{k_x k_y k_z} \qquad (7)$$

$$k_v = k_z = 0.0718 \times \sqrt{\left[\frac{k_H(1-V_{sh})}{\phi_e}\right]^{2.0901}} \qquad (8)$$

The results of the permeability anisotropy calculation are given in Table 3.

Table 3. Permeability anisotropy near the well 16/2-D-12.

Parameter	k_x	k_y	k_z	a
Value	22.2 D	22.2 D	6.4 D	0.3

The wettability in a reservoir depends on several factors like polar components in oil, formation water composition, and salinity as well as rock composition. By considering all these parameters and criteria it can be believed that the Johan Sverdrup reservoir has a mixed-wet wettability state (Fang, 2017).

Since there is no information about the relative permeability of the Johan Sverdrup reservoir, this parameter must be estimated based on the available information about the relative permeability of the other sandstone reservoirs in the North Sea with almost the same characteristics. The Frøy field is located in the central part of the North Sea. Frøy is a sandstone reservoir with high porosity and absolute permeability almost like the Johan Sverdrup reservoir. It also has some mixed-wet zones like the Johan Sverdrup reservoir. As a result, the values of relative permeability of the Johan Sverdrup reservoir can be considered almost the same as that of the Frøy reservoir in the mixed-wet zones (Tangen, 2012; Hadia et al, 2011). Figure 4 shows the relative permeability curves based on the relative permeability in the mixed-wet zones of the Frøy field and they can be used for the Johan Sverdrup reservoir as well.
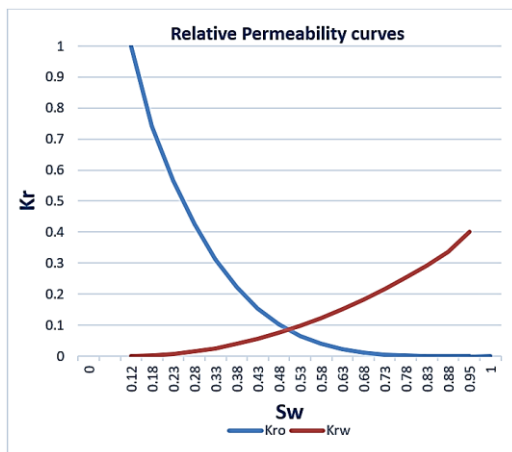
Figure 4. Relative permeability curves for the Johan Sverdrup reservoir.

4 Development of the OLGA/ROCX model

The combination of OLGA and ROCX creates one of the leading and robust tools for modeling and simulation of multiphase flow behavior from the reservoir pore to the production pipe and process facilities. OLGA is a dynamic multiphase flow simulator and ROCX is a reservoir simulator that can be coupled to OLGA as a plug-in.

4.1 Development of the reservoir model in ROCX

In this subchapter, the main settings for developing the model of the reservoir near the well 16/2-D-12 in ROCX are described.

4.1.1 Determining the dimensions of the reservoir drainage area and the grid setting

In order to create a near-wellbore model of oil production, the first step is to determine the dimensions of the drainage area near the well. The drainage area of a horizontal well has an ellipsoidal shape. However, due to the ROCX limitation for creating an ellipsoidal geometry, a rectangular drainage area is considered for developing the model. As mentioned in Table 2 the total thickness of the net pay reservoir near the well 16/2-D-12 is 55.5 m (32.46 + 22.86 = 55.32 ~ 55.5 m). Therefore, the height of the drainage area is considered to be equal to 55.5 m.

There is no information about the exact geometry of the well 16/2-D-12 in the literature. Therefore, based on the classification of horizontal wells it is assumed that the well 16/2-D-12 is a medium radius horizontal well with the maximum radius in the kickoff section which means $R_{kickoff} = 305 \, \text{m}$. According to the NPD's fact pages (Norwegian Petroleum Directorate, 2020), the measured depth of the well 16/2-D-12 is $L_{MD} = 3875$ m and it has the final vertical depth of $L_{TVD} = 1876$ m. As a result, the length of the horizontal section of this well, $L_{horizontal}$, can be calculated by Equation 9.

$$L_{MD} = L_{TVD} + L_{kickoff} + L_{horizontal} \tag{9}$$

By using Equation 9 the length of the horizontal section of the well is calculated as 1612 m. Therefore, the length of the drainage area is considered the same as that of the horizontal section of the well. To determine the width of the drainage area, oil production from five test cases with similar heigh and length but different width of the drainage area is simulated in OLGA. The simulations are conducted for a horizontal well with a length of 124 m and one ICD valve in a reservoir with the same rock and fluid properties as the JSF. Figure 5 shows the accumulated oil production during 120 days of simulation for the given cases. As can be seen in the figure, by increasing the width of the drainage area, the time of water breakthrough and accumulated oil production increases, but it has a converging pattern in such a way that there is no big difference between accumulated oil production in the cases with a width of 120 m and 140 m. As a result, it can be concluded that

considering the width of the drainage area for the main model equal to 120 m can be a good estimation. Of course, considering the width of the drainage area equal to 140 m leads to more accurate results, however, it also increases the time of the simulation.
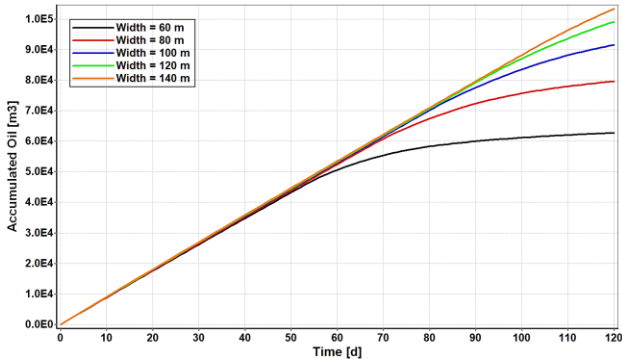
Figure 5. Analysis of the sensitivity of oil production to the width of the drainage area.

Since in the JSF oil is produced by water drive, the well must be positioned as near as possible to the top of the drainage area to delay water breakthrough. Therefore, for developing the model, it is assumed that the well is located 5.5 m below the top of the drainage area. The geometry of the drainage area and the position of the well, considered for developing the model of oil production from the well 16/2-D-12 in the JSF is schematically depicted in Figure 6.
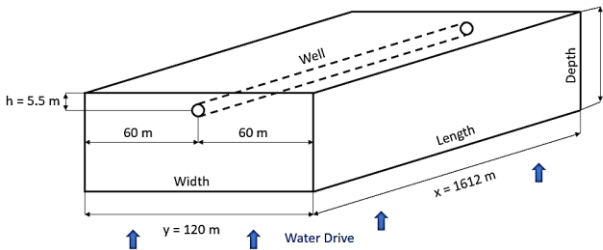
Figure 6: Schematic geometry of the drainage area.

To discretize the reservoir in ROCX, the number of grids in (x,y,z) coordinates must be determined. In general, a suitable grid setting can be achieved by using finer mesh in the places with high variation in the fluid properties and coarser mesh in the other places. Since the cross-section of the reservoir is located in the Y-Z plane, the fluid pressure in the Y-Z plane around the well drops significantly. Therefore, to achieve more accurate results, finer mesh in the y and z directions close to the well must be defined in the grid setup. The length of the well is in the x-direction. As a result, the fluid pressure has small variations in the x-direction, and considering uniform mesh in the x-direction can maintain enough accuracy. In order to develop the model, it is assumed that the well 16/2-D-12 contains 13 joints, each 124 m long and consisting of one equivalent valve. As a result, 13 uniform cells are considered for the reservoir in the x-direction. In the y and z directions,

the meshes are not uniform, and the grid resolution in these directions is shown in Figure 7.
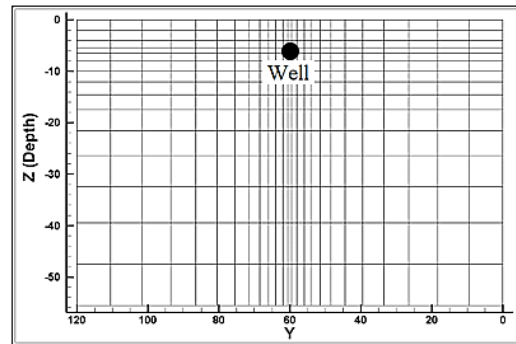
Figure 7. Grid resolution in the Y-Z plane.

4.1.2 Reservoir property settings

Since the Johan Sverdrup reservoir is heterogeneous with respect to permeability, for the investigation of the oil production from the well 16/2-D-12, the OLGA/ROCX model is developed for a heterogeneous reservoir. The permeability in the JSF varies between 0.5 and 40 D while the average permeability in the reservoir near the well 16/2-D-12 is 14.7 D. As a result, there are some places in the reservoir where the permeability is more than the average permeability. Moreover, heterogeneity may exist in the horizontal, vertical, or angled direction. Therefore, for creating a heterogeneous case, the permeability in some random zones in the horizontal, vertical, and angled directions is considered 1.5, 2, and 2.5 times higher than the average permeability in the near-well reservoir. MATLAB is used for generating required data for the implementation of heterogeneity in ROCX. By considering the permeability anisotropy given in Table 3, the average permeability in the x and y directions is equal to 22.2 D, and average permeability in the z-direction is equal to 6.4 D throughout the reservoir. The permeability distribution in the heterogeneous reservoir in the z-direction used for developing the OLGA/ROCX model is shown in Figure 8. The permeability distribution in the x and y directions has the same pattern.
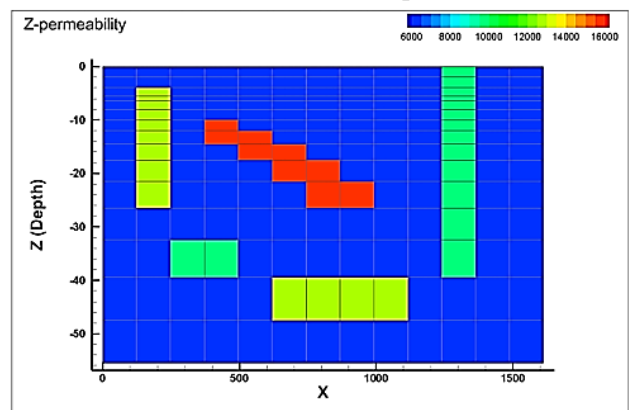
Figure 8. Permeability distribution in the reservoir.

4.1.3 Initial condition settings

Based on the information given in Table 2, the reservoir near the well 16/2-D-12 consists of two zones with different types of sandstone. Therefore, for developing the model it is assumed that initially the reservoir is filled with oil in both zones, but with a water saturation of 0.062 and 0.322 in the first and second zone respectively. Figure 9 represents the initial water saturation in the reservoir.
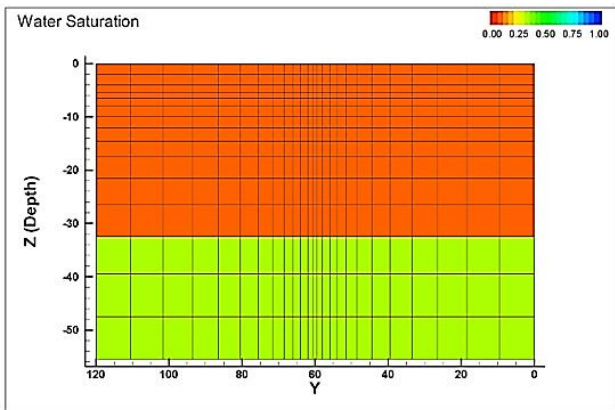
Figure 9. Initial water saturation in the reservoir.

4.2 Development of the well model in OLGA

For developing the well model in OLGA, one pipe with a length of 1612 m, a diameter of 0.1397 m (5.5 inch), and roughness of 15 μm is considered for representing the production tubing. Another pipe with the same length but a diameter of 0.2159 m (8.5 inch) is considered for representing the wellbore. It is assumed that oil is produced from 13 zones in the well and each zone contains two hypothetical sections. Therefore, the production tubing and wellbore are discretized to 26 hypothetical sections, each 62 m long. The simplified model for oil production from each zone is represented in Figure 10.
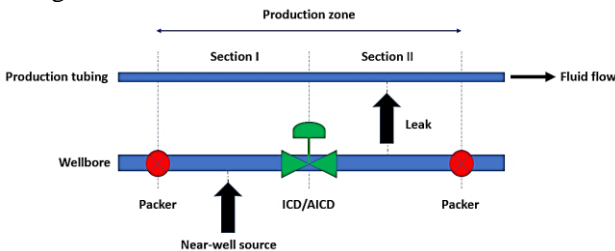
Figure 10. Simplified representation of a single production zone.

As can be seen in the figure, in order to stop flowing the reservoir fluids between different zones in the annulus, each production zone in the wellbore is separated by two packers. Moreover, by using a near-well source connected to the ROCX model, the reservoir fluids enter the wellbore after passing an inflow control device located in the first section of the wellbore. The reservoir fluids that pass through the inflow control device enter the production tubing via a leak connected to the second section of the production tubing and move towards the heel of the well. This setup has been proposed and used in (Aakre, 2017).

Based on the estimation of oil production from the well 16/2-D-12, and considering the frictional pressure drop in the well and pressure difference across the ICDs and AICDs, the pressure drawdown for this well is considered to be 12 bar. Moreover, the hole diameter of the equivalent valve is calculated as d = 0.0108 m. In order to implement the autonomous behavior of the AICD in OLGA, a controller must be used for chocking the AICD based on the characteristics of the AICD and reservoir fluid mixtures. For developing the model for oil production from the well 16/2-D-12 with RCP valve completion, the *Table Control* module in OLGA is used for controlling the valve. In this model, at first, by using the *Transmitter* module in OLGA, the water cut of the fluid mixture is measured. Then the Table Control, based on the measured water cut of the fluid mixture and some tabulated data (based on experimental data for the behavior of the RCP valve), provides a corresponding control signal for partially chocking the RCP valve.

5 Results and discussion

In this chapter, the obtained simulation results from the OLGA/ROCX model are presented and discussed. Besides, the functionality of ICD and RCP in improving oil recovery is evaluated. The simulations have been conducted for three cases: a) open-hole well, b) well with ICD completion, c) well with RCP completion.

5.1 Fluid flow distribution and time of water breakthrough

Figure 11 shows the oil saturation contours near the open-hole well right after the water breakthrough. The well consists of 13 zones and the 3D contour illustrates the oil saturation distribution from all the 13 zones.
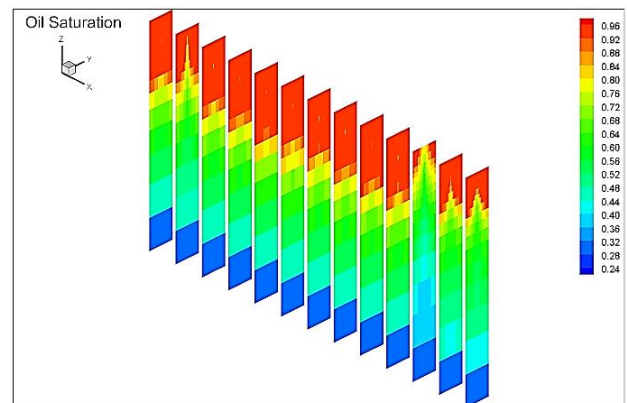
Figure 11. Oil saturation distribution right after the water breakthrough in the open-hole well.

As shown in the contours, once the water breakthrough takes place, the water saturation profile has a conical shape. Besides, it can be clearly seen that the water cone grows faster and breaks into the well sooner at the heel compared to the toe of the well. It is obvious that the early water breakthrough occurs in the high permeable zones (the 2nd and 11th zones) and the water cone develops much faster in these zones. Based on the obtained results, the water breakthrough for the open-hole well takes place on the 3rd day of production.

The oil saturation contours for the reservoir near the well 16/2-D-12 with considering inflow control devices right after the water breakthrough are illustrated in Figure 12. The well consists of 13 zones and the 3D contour illustrates the oil saturation distribution from all the 13 zones.

Figure 12. Oil saturation distribution right after the water breakthrough for the well with inflow control devices.
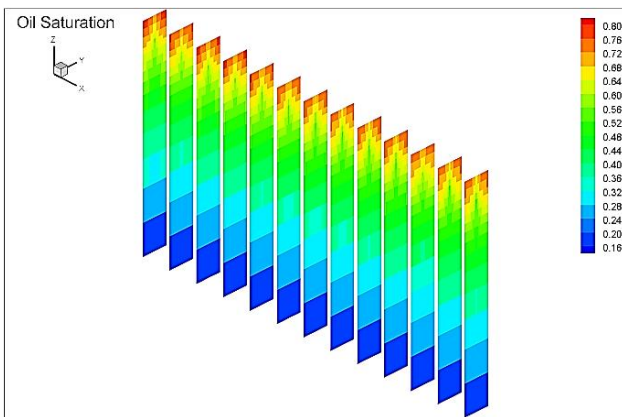
The simulation results show that the water breakthrough takes place on the 258th day of production. As a result, by completion of the well 16/2-D-12 with ICD or RCP valves, the water breakthrough can be delayed by 255 days compared to the open-hole well. Moreover, By comparing Figure 11 and Figure 12, it can be concluded that the inflow along the well completed with inflow control devices is much more balanced compare to the open-hole well. Therefore, by using the inflow control devices the flow influx along a horizontal well can be effectively evened out.

5.2 Accumulated oil and water production

In order to investigate oil production and comparing the performances of different inflow control devices, accumulated oil and water are the two most important parameters that must be taken into account. Figure 13 illustrates the accumulated oil and water produced from the well 16/2-D-12 by considering ICD and RCP completions. As can be seen in the figure, after 750 days of production, there is a very small difference between accumulated oil in the ICD and RCP cases. However, due to the choking effect of the RCP valve for low viscous fluids, the amount of accumulated water drops by 11.9 % when the well is completed with RCP valves.
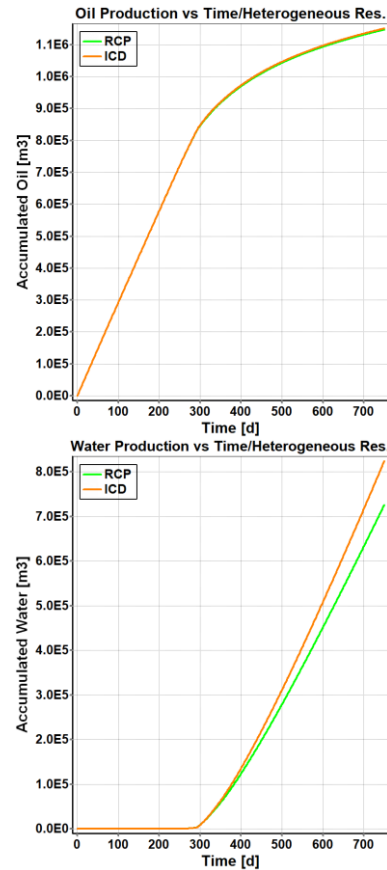
Figure 13. Accumulated oil and water from the well 16/2-D-12 with ICD and RCP completions.

5.3 Oil and water flow rate

Figure 14 illustrates the volumetric flow rate of oil and water production from the well 16/2-D-12. As shown in the figure, just after starting the production, the oil flow rate in the RCP and ICD cases reaches its peak. By increasing the water saturation around the wellbore in the subsequent days, the rate of oil production experiences a slight decrease until the time of water breakthrough. At the time of water breakthrough, the water saturation near the wellbore exceeds the irreducible water saturation and as a result, water enters the well. Once water begins to enter the well, the reservoir tends to produce more water than oil. Consequently, after the water breakthrough, the rate of oil production drops substantially. The oil production can be continued until the oil saturation falls below the residual oil saturation. Therefore, after 750 days some amount of oil is produced. According to Figure 14, over the whole period of production (750 days), oil production in both the ICD and RCP cases is almost similar. As a result, it can be concluded that practically there is no difference between using ICD and RCP with respect to oil flow rate over the first 750 days of production. However, after water breakthrough, due to the capability of the RCP valve for choking the water back, the diagram of the water flow rate in the RCP case, falls below that of the ICD case. The simulation results

show that the well completed by RCP valves reduces the rate of water production by 13.4% compared to using ICDs after 750 days of production.
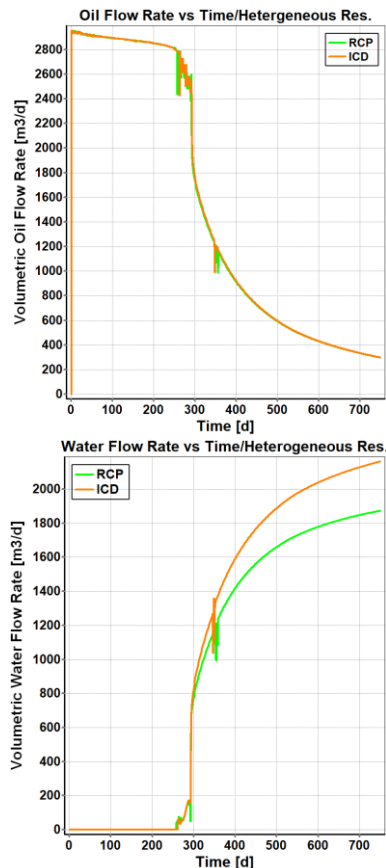
Figure 14. Flow rate of oil and water for the well 16/2-D-12 with ICD and RCP completions.

6 Conclusion

The obtained results show that for both the open-hole well and the well with inflow control devices, water begins to enter the well from the heel side due to the heel-toe effect. Moreover, due to the faster development of the water cone in the high permeable zones, the early water breakthrough takes place sooner in these zones. Early water breakthrough leads to a significant drop in oil production, and this problem must be tackled to achieve cost-effective oil production. Both ICDs and RCPs can effectively even out the flow influx along the horizontal well and consequently delay the water breakthrough significantly. Since RCPs remain fully open before the time that water breakthrough occurs, there is no difference between the functionality of ICDs and RCP valves in delaying the time of water breakthrough. However, since the RCP valve opening is sensitive to the fluid viscosity, these valves have a capability for chocking the unwanted fluids. According to the simulation results, using RCP valves for the JSF condition can significantly reduce water production. By completion of the well 16/2-D-12 with RCP valves, the accumulated water production can be reduced by 11.9%

during the first 750 days of production. Besides, using RCP valves can reduce the flow rate of water production by 13.4 after 750 days. Furthermore, by evaluating the simulation results it can be concluded that using RCP valves has a negligible impact on both the accumulated oil production and flow rate of oil production from this well compared to using ICDs. Therefore, by completion of the well 16/2-D-12 with RCP valves more cost-effective oil production can be achieved.

Refferences

Haavard Aakre. *The impact of autonomous inflow control valve on increased oil production and recovery*. PhD thesis, University College of Southeast Norway, 2017.

Stian M. Askvik and Ivar L. J. Sørheim. *Dynamic Autonomous Inflow Control Device-Performance prediction and experimental investigation of a specific rate controlled production valve design*. Master's thesis, NTNU, 2017.

Omar J. Chammout, Bisweswar Ghosh and Mohamad Y. Alklih. Downhole flow controllers in mitigating challenges of long reach horizontal wells: A practical outlook with case studies. *Journal of Petroleum and Gas Engineering*: 97-110, 2017.

Equinor. Crude oil assays. 2020. https://www.equinor.com/no/what-we-do/crude-oil-and-condensate-assays.html.

Wei Fang. *Evaluation of low salinity injection EOR potential in Johan Sverdrup Field*. Master's thesis, University of Stavanger, 2017.

Nanji Hadia, Havard H. Lehne, Kanwar G. Kumar, Kristoffer A. Selboe, Feb Å. Stensen and Ole Torsater. Laboratory Investigation of Low Salinity Waterflooding on Reservoir Rock Samples from the Frøy Field. *SPE Middle East Oil and Gas Show and Conference, Bahrain, 2011*.

Martin Halvorsen, Martin Madsen, Mathias V. Mo, Ismail I. Mohd, and Annabel Green. Enhanced oil recovery on troll field by implementing autonomous inflow control device. *SPE Bergen One Day Seminar, Norwey, 2016*.

Philip C. Iheanacho, Djebbar Tiab and Alpheus O. Igbokoyi. Vertical-Horizontal permeability relationships for sandstone reservoirs. *Nigeria annual international conference and exhibition, Nigeria, 2012*.

Hans-Martin Kaspersen. *Reservoir Characterization of Jurassic Sandstones of the Johan Sverdrup Field, North Sea*. Master's thesis, University of Oslo, 2016.

Vidar Mathiesen, Bjernar Werswick, Haavard Aakre and Geir Elseth. Autonomous Valve, A Game Changer Of Inflow Control In Horizontal Wells. *Offshore Europe Oil and Gas Conference and Exhibition, UK, 2011*.

Norwegian Petroleum Directorate. Fact pages. 2020. https://factpages.npd.no/en/field/pageview/all/26376286.

Mathias Tangen. *Wettability Variations within the North Sea Oil Field Frøy*. Master's thesis, NTNU, 2012.

The Engineering Toolbox. Orifice, Nozzle and Venturi Flow Rate Meters. 2004. file:///C:/Users/alimo/OneDrive/Skrivebord/Orifice/Orifice,%20Nozzle%20and%20Venturi%20Flow%20Rate%20Meters.html.

Sensitivity Analysis and Case Studies for CO₂ Transportation Energy Consumption.

Steven Jackson

Department for Engineering & Technology (IVT), UiT – Norway's Arctic University, Norway,
`steve.jackson@uit.no`

Abstract

The transportation of CO_2 is important to all carbon capture and storage (CCS) projects. Both the infrastructure costs (compressors, pipelines, tanker ships, etc.) and the energy consumed in the compression or liquefaction of CO_2 are significant. Understanding how the size, capacity and energy consumption of transportation alternatives varies between projects is therefore important. Modelling provides a useful insight into the performance of transportation alternatives, but the results are only useful when the basis for comparison is consistent and the impact of model input parameters is well understood. This article presents the results of sensitivity studies made using a transportation model that was developed in earlier work. Several important model parameters are studied using three planned/operating CCS project cases. The results show that while the operating pressure of the storage site is most important in determining the transportation system operating pressure, the temperature of the available cooling utility is the key parameter determining energy consumption.

Keywords: CO₂, CCS, transportation, modelling

1   Introduction

All carbon, capture and storage (CCS) projects require the transportation of CO_2 from a source to a storage location. A compressor and a large diameter pipeline is the method often used to achieve this, but as illustrated in Figure 1, the liquefaction of CO_2 to allow ship–based transportation can also form one of the links in the transportation process.

Although identifying the optimum economic case is of key importance to all CCS projects, it is also important to minimize energy consumption because the energy consumed by the process corresponds directly to the efficient consumption of non-renewable resources in fossil-fuel based CCS projects.

Most of the energy consumption associated with CO_2 transportation comprises compression and pumping energy. Compressors are often used to raise the pressure of gaseous CO_2 streams or gaseous refrigerants (in the case of liquefaction) and pumps are used to raise the pressure of liquid CO_2 streams. The pressure-level required for transportation depends on the operating parameters of the storage location, the design of the pipeline and the temperature under which the pipeline operates. Understanding how the combined effect of these parameters affects energy consumption can, therefore, provide an important insight into the relative strengths of different CCS projects.

As part of a project aimed at studying the performance of CCS project alternatives a MATLAB based model for the transportation of CO_2, *CO2TM,* has been developed and is made freely available at UiT Open Research Data (Jackson, 2020a).This model is used as the basis for the present study.

The *CO2TM* takes inputs comprising the source location, transportation type (ship or pipeline) pipeline route, storage reservoir depth and CO_2 mixture type. From these inputs, the model calculates an elevation profile for pipelines and a temperature profile using built-in bathymetry (seabed elevation profile) and sea surface temperature (SST) data. Based on elevation, temperature and CO_2 mixture data, the model estimates the reservoir and wellhead pressure (WHP) and then determines the pressure profile required to ensure
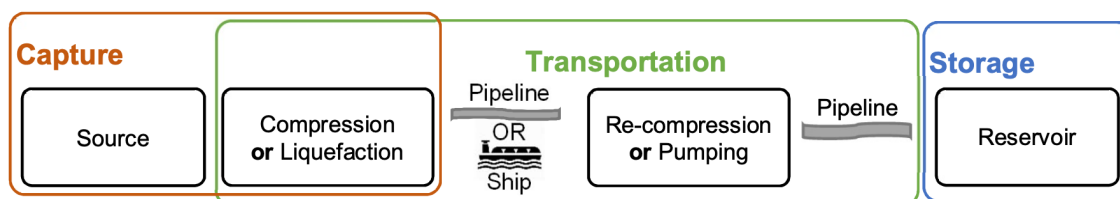
Figure 1. Illustration of the Main Associated with the Transportation of CO₂ (from Jackson, 2020).

single-phase flow in the pipeline. Finally, the energy consumption for transportation processes—comprising either compression or liquefaction—is calculated using the ambient sea temperature in location where each part of the process is situated. The development of this model is described in detail in earlier work (Jackson, 2020b).

The aim of this article is to present results from a study into the sensitivity of the *CO2TM* to various modelling parameters so that the application of this model can be better understood. The study is roughly based on three planned/ operating CCS project cases, which are used to illustrate the impact of the studied parameters on performance.

2 Method

Because the focus of earlier work—including the development of the *CO2TM*— has been the impact of ambient temperature on CO_2 transportation system performance, the main focus for this work is also the study of model sensitivity to seawater temperature. A related modelling parameter, also studied in this work, is the pipeline Heat Transfer Coefficient, HTC. In addition, this article presents results for CO_2 transportation system sensitivity to pipeline roughness, mixture composition and transportation type. The method used in the study of each of these parameters is described in more detail below under several sub-headings.

Although the development of the *CO2TM* is described in earlier work, some modifications to the original model were required to facilitate the present study. The modifications made are also described under the headings set-out below where they are relevant and will be subsequently included in an updated version of the *CO2TM*.

2.1 Sensitivity to Temperature

The *CO2TM* determines SST in the locations defined in the model input parameters using data from JMA (Japan Meteorological Agency). The resulting temperature data is then used as the basis for calculation of the energy consumption of the liquefaction and compression processes along with the temperature profile in the pipeline. For the compression and liquefaction processes, the model applies a margin of 10 °C above the seawater temperature.

Because SST varies annually, the data used in the model is based on two standard deviations above the average of yearly SST, i.e. covering around 95% of all SST measurements. This results in a conservative estimate for the energy consumption of compression and liquefaction processes and the pipeline temperature profile. It is therefore natural to study the sensitivity of the model to seawater temperature with an emphasis on reduced temperatures, which can be interpreted as either the performance during winter months or a less conservative approach to heat exchanger design. To

reflect this, a range of temperatures from base -8 °C to base +4 °C is used in the sensitivity studies conducted here.

To allow the study of this temperature range, the original *CO2TM* required some modification. The main modification was to allow the user input of seawater temperature to apply to both the liquefaction location and the pipeline location for cases where transport is by ship. This represents an over simplification of reality where the liquefaction location may be a significant distance from the storage location, but it is one way in which sensitivity can be studied.

Another modification required was to implement limits on the minimum sea temperature used in the code to avoid extrapolation of parameters such as density and heat capacity outside of the range of the basis data included in the model. This was done by setting a minimum possible SST of 5 °C within the model.

2.2 Sensitivity to Pipeline Roughness

Pipeline roughness affects pipeline pressure-drop and can vary with both construction material and the age of the pipeline, equating to corrosion and fouling over time. In large diameter gas pipelines a coating is sometimes used to reduce pressure-loss and studies have found that absolute roughness can be as low as 4 μm (Langelandsvik, 2008). However, studies relating to CO_2 pipelines have often used higher values of roughness ranging up to 100 μm (Chandel et al., 2010). The default value of roughness used in the model is 15 μm, but this can be over-ridden using a user-specified roughness input parameter. In the present study, the roughness input parameter was varied from 2 to 100 μm to provide a range of results illustrating sensitivity.

2.3 Sensitivity to Heat Transfer

Subsea pipelines typically loose heat along their length to the surrounding seawater. The HTC, which varies with pipeline design and burial conditions, determines how quickly the pipeline contents approaches the sea temperature. In-turn, the temperature in the pipeline can impact the required operating pressure, which must be maintained at a margin above the bubble point curve of the CO_2 mixture throughout the pipeline.

The default value of the coefficient used in the model is 4 W/m^2 K, but the user can override this using a user-specified model input parameter. Studies of onshore buried pipelines have used HTC in the range 1–6 W/m^2 K (Mazzoccoli et al., 2014; Zhang et al., Massarotto et al., 2006), and for pipelines surrounded by water, up to 45 W/m^2 K (Drescher et al., 2013). The present study uses a range from 1 to 32 W/m^2 K to investigate the impact of this parameter on CO_2 pipelines.

2.4 Sensitivity to CO$_2$ Mixture Composition

The composition of CO_2 mixtures in transport systems depends on the source of the CO_2 and the entry

specifications set for the transportation system. The *CO2TM* has three built-in CO_2 mixture compositions with associated property data representing post, pre and oxyfuel combustion CO_2 sources.

In the previously published version of the *CO2TM,* the post combustion case is the only mixture composition made available for use. To enable the study of the sensitivity to CO_2 mixture composition in the present work, an update was required to make the oxyfuel mixture composition available for use. This work was done on the same basis as the earlier work and, although the details of the method are not described here, the composition used is provide in Table 1.

Table 1. CO_2 Mixture Compositions.

Component	Post	Oxyfuel
CO_2 mole %	99.99	96.16
N_2 mole %	0.01	2.45
Ar mole %	–	0.96
O_2 mole %	–	0.43

In addition, to provide a consistent basis for comparing energy consumption between post combustion and oxyfuel cases, an update of the *CO2TM* was required to allow the liquefaction energy consumption to be calculated for cases where the feed stream has a pressure of 15 barg— e.g. originating from a low–temperature type oxyfuel purification unit. The reduction in liquefaction energy for these cases was estimated by taking the difference between the compression energy for pipeline transport for the oxyfuel and post combustion capture cases and then deducting this from the energy consumption of the standard liquefaction process, where the feed stream is at low pressure. The updated version of the *CO2TM* will be published subsequent to the completion of the present work.

The sensitivity study conducted in the present work is based on a comparison of the performance of post and oxfuel combustion CO_2 mixture compositions. The basis of this comparison is both the transportation energy consumption and the pipeline inlet pressure. Results are summarized for the *CO2TM* default pipeline size selection: the first pipeline size that results in a pipeline pressure under 180 barg for all operating cases, and for the case where all pipelines have the same diameter.

2.5 Sensitivity to Transportation Method

The transportation cases used in this work are loosely based on three planned/operating CCS projects. Case 1 reflects the planned Norcem/Northern Lights (NL) project , which includes ship-based transport of CO_2

from Norcem in Brevik to the planned NL storage hub in south east Norway. Case 2 reflects the proposed H21 project, which is planned to include the conversion of natural gas to hydrogen with carbon capture in the UK with CO_2 storage in the North Sea[2]. Case 3 reflects the Melkøya CCS project, where CO_2 is removed from natural gas and returned to storage in the Barents Sea. A summary of some of the main modelling parameters associated these cases is provided below in Table 2.

Table 2. Comparison of Case Parameters.

Parameter	Case 1	Case 2	Case 3
Source Location	9,69 E 59,06 N	0,12 E 53,65 N	23,59 E 70,69 N
Liquefaction Loc.	9,69 E 59,06 N	-	-
Compression Loc.	-	0,12 E 53,65 N	23,59 E 70,69 N
Pipeline location	4,89 E 60,56 N	0,12 E 53,65 N	23,59 E 70,69 N
Pipeline length (km)	107	129	151
Reservoir location	3,42 E 60,45 N	2,00 E 54,00 N	4,89 E 60,56 N
Wellhead depth (m)	300	76	318
Reservoir Depth (m)	2000	1300	2500
Sea Temp. (°C)*	15,3	18,0	10,9

* Calculated by the *CO2TM*

In addition to the three cases described above, three alternative cases are also defined: Case 1A is the NL project with pipeline transport of CO_2 directly from the pipeline location; Cases 2A and 3A reflect Cases 2 and 3 with shipping to the NL pipeline as an alternative to pipeline transportation.

The model parameters used to specify the pipeline route for all cases and reservoir details are inferred from openly available data and should not be taken to accurately reflect the details of these projects. Figure 2 provides an illustration of the pipeline route used for the NL cases that was generated using the *CO2TM*.

3 Results & Discussion

The main results of the study are set out below under separate sub-headings.

3.1 Sensitivity to Pipeline Heat Loss

Figure 3 presents results that illustrate the sensitivity of pipeline inlet pressure to the heat transfer coefficient used in the model. For Cases 1A, 2 and 3 a small impact on pressure is visible, but in Case 1 there is almost no impact. This can be explained by the fact that in Case 1

the CO_2 mixture is very close to the seawater temperature at the point of entry to the pipeline.
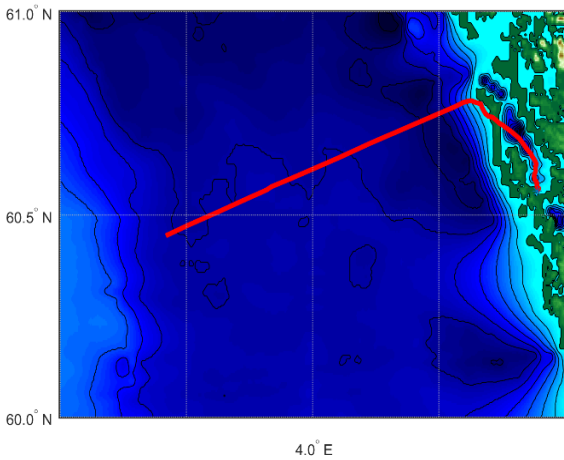
Figure 2. Illustration generated by the *CO2TM* for the Norther Lights (NL) pipeline route used in this work.
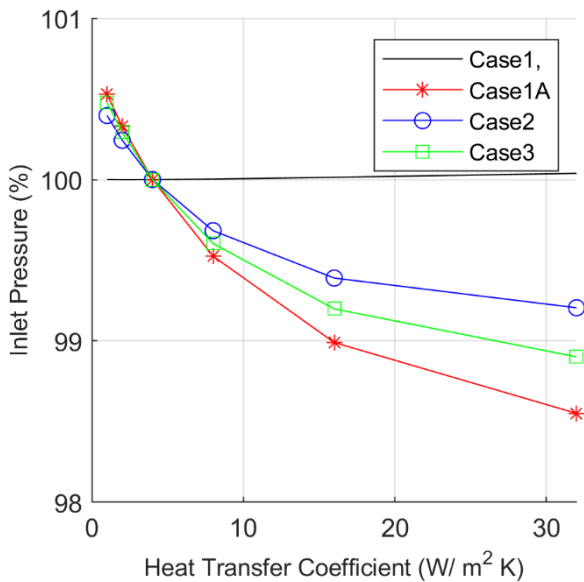
Figure 3. Variation in Pipeline Inlet Pressure with HTC where 100% is the model default basis of 4 W/m^2 K.

Figure 4 shows that the variation in pipeline inlet pressure presented in Figure 3 equates to an even smaller variation in overall energy consumption, reflecting the fact that the dominant part of the system energy consumption is associated with the earlier stages of compression, in the compression cases, and with the liquefaction process in the liquefaction cases.

3.2 Sensitivity to Pipeline Roughness

Figure 5 and Figure 6 present results for the sensitivity of pipeline inlet pressure and energy consumption to pipeline roughness. They show that roughness is a more important factor in transport system design than the HTC, although Figure 6, like Figure 4, shows that the roughness does not play a big role in determining the system energy consumption.
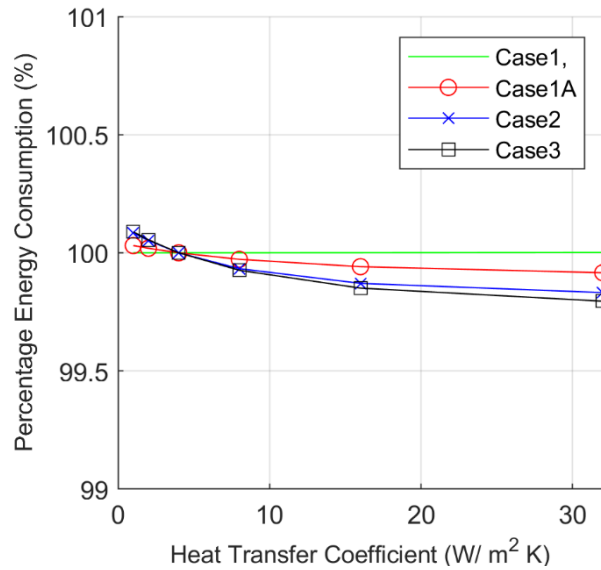
Figure 4. Variation in Energy Consumption with HTC.

However, Figure 5 does show that roughness can have a significant impact on the pipeline operating pressure, which is important to selection of the pipeline size and therefore the economics of CCS projects.
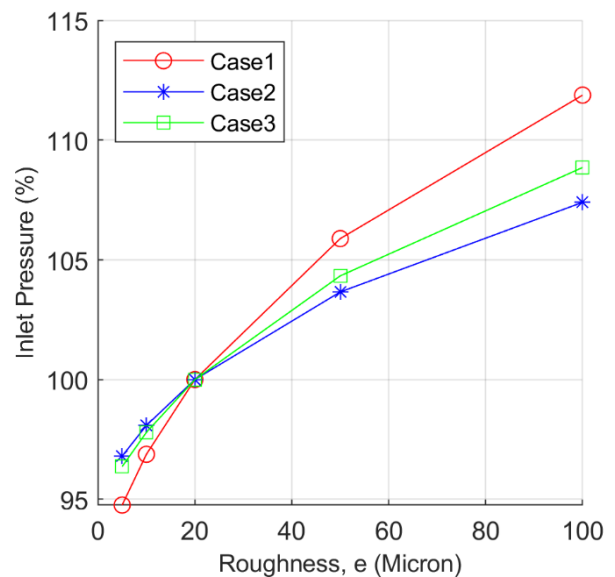
Figure 5. Variation in Percentage Pipeline Inlet Pressure with Pipeline Roughness.

3.3 Sensitivity to Temperature

Figure 7 and Figure 8 show the impact of seawater temperature on pipeline inlet pressure and energy consumption when the default seawater temperature estimated by the model is adjusted in the range – 8 ºC to + 4 ºC. The results are split into cases with pipeline-based transport (shown in red) and shipping based transport (shown in blue).
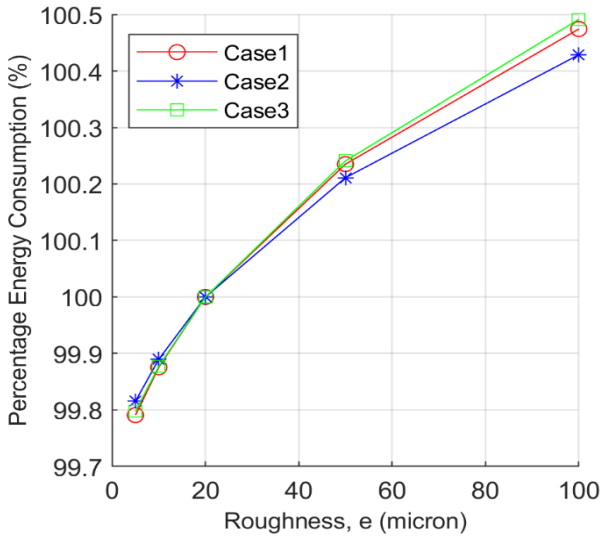
Figure 6. Variation in Percentage Energy Consumption with Pipeline Roughness.

The results show that although the impact of SST on pipeline inlet pressure is not more pronounced than that of roughness, the impact on energy consumption in all cases is much more significant. This is due to the dual impact of sea temperature, i.e. that it both affects the pipeline operating pressure and reduces the energy consumption of the associated compression and liquefaction processes.

The results presented for Case 3, Melkøya in Figure 7 and Figure 8 also show that when the seawater temperature is reduced by 8 °C, the temperature of the pipeline falls below the lowest temperature where compressor energy consumption data is available in the model. In reality, there would be some continued reduction in energy consumption that would gradually reduce towards zero as the sea temperature is also reduced towards zero.
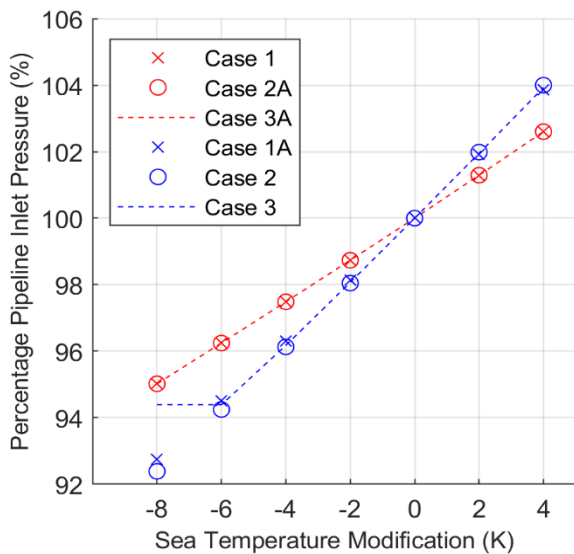
Figure 7. Variation in Energy Consumption with Seawater Temperature for Shipping Cases (in red) and Pipeline Transport (in blue).
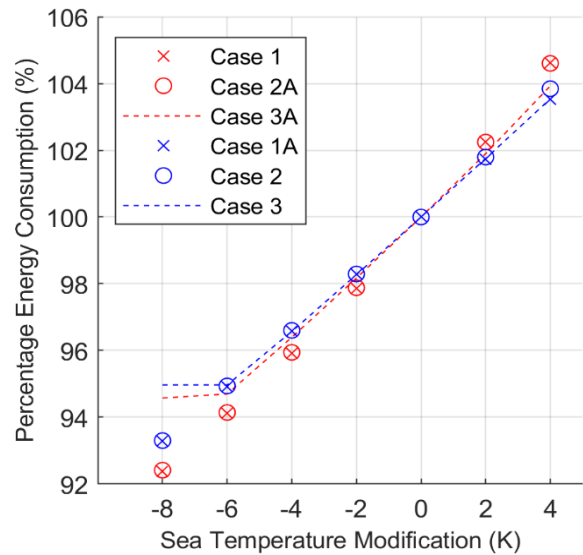
Figure 8. Variation in Energy Consumption with Sea Temperature for Shipping Cases (in red) and Pipeline Transport (in blue)

3.4 Sensitivity to CO₂ Mixture

Figure 9 shows how the pipeline pressure varies with sea temperature for two CO_2 mixture compositions representing post combustion capture and oxyfuel combustion.

All of the results indicate a small increase in operating pressure for the oxyfuel cases. This is due to an increased CO_2 mixture bubble-point pressure, which affects the minimum pipeline operating pressure: in all cases the *CO2TM* enforces a margin between bubble-point pressure and operating pressure. The sensitivity of inlet pressure to seawater temperature is similar for most cases.

Figure 9 shows results for the Norcem/NL case on two different pipeline design basis: a 14-inch pipeline sized based on the model default parameter of 180 bar maximum pipeline operating pressure (red lines), and a 16-inch pipeline sized to match the other two cases (black lines). This comparison highlights an inherent advantage of the NL pipeline that results from a combination of wellhead depth and reservoir depth (see Table 2). The results show that both of these factors have an important influence on the pressure profile calculated by the *CO2TM*.

How the variation in Inlet Pressure translates into a variation in energy consumption is presented for the post combustion cases, which are discussed under the next heading.

3.5 Sensitivity to Transportation Type

Figure 10 shows the variation in energy consumption with seawater temperature modification for a selection of pipeline and ship-based transportation cases.

The results show that in all of these cases, ship based transportation consumes more energy than sending the
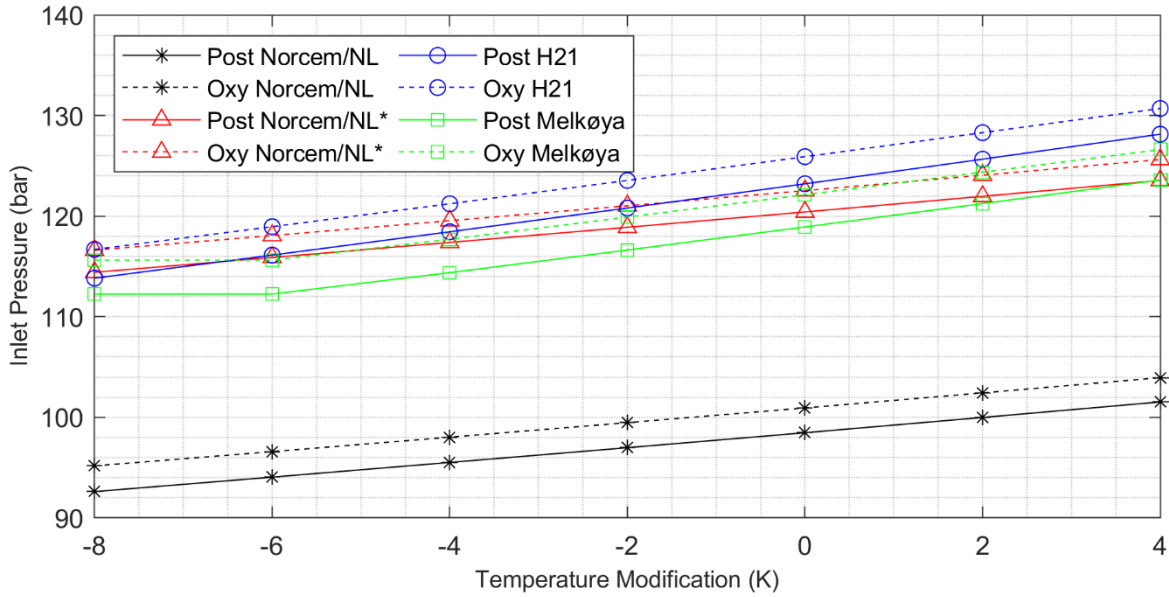
Figure 9. Impact of CO_2 Mixture Composition on Pipeline Operating Pressure.

CO_2 to a pipeline. Although the proportion of additional energy required is seen to vary between cases, the sensitivity of energy consumption to seawater temperature is similar for all cases.

Similar to Figure 9, Figure 10 presents results for both a 14-inch and a 16-inch NL pipeline diameter. However, the results presented in Figure 10 show that the impact of increasing the pipeline size on energy consumption is much smaller than the impact on pipeline operating pressure.

The results presented in Figure 10 also show that, regardless of transportation type, the energy consumption varies significantly between cases: both pipeline transport and shipping CO_2 from Melkøya results in the lowest energy consumption of all cases.

This highlights the important role that ambient temperature plays in determining the energy consumption for both transportation alternatives.

4 Conclusions

When pipelines operate close to the temperature of the surrounding medium, the heat transfer coefficient has a low impact on operating pressure and energy consumption.

Pipeline roughness has a small, but potentially important impact on CO_2 pipeline operating pressure and hence the selection of the economic optimum pipeline diameter. The impact of roughness on the
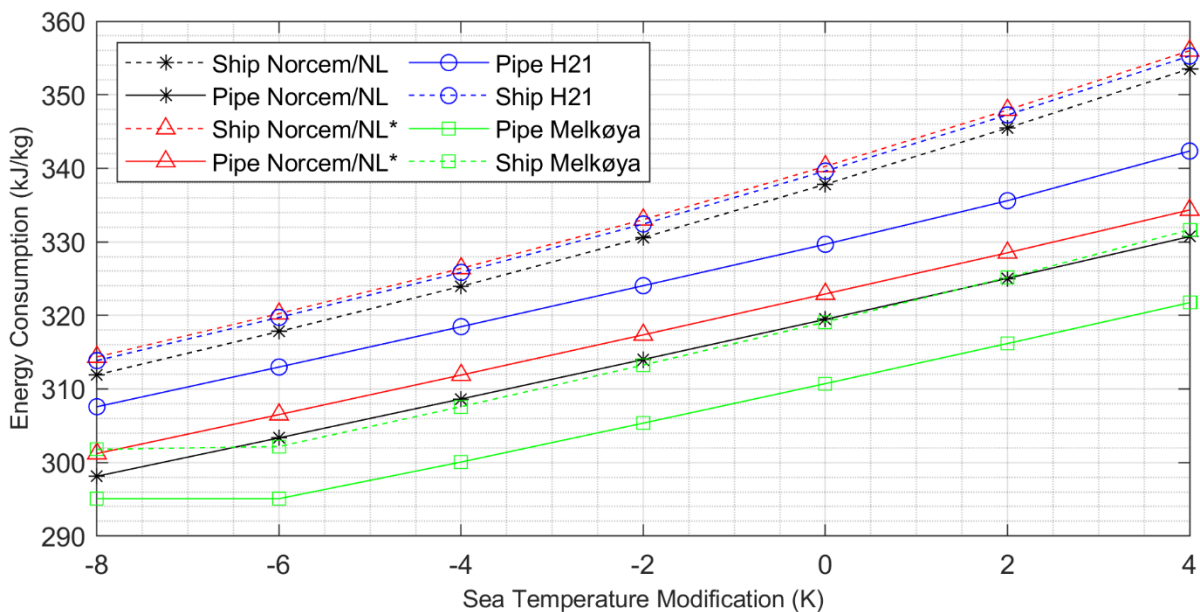
Figure 10. Impact of Transportation Type on Energy Consumption with Varying Sea Temperature

energy consumption associated with transportation of CO_2 is, however, small.

The composition of the CO_2 mixture transported in a pipeline can have an important impact on the transportation pressure.

The operating pressure of the storage reservoir and the wellhead location have the most important impact on CO_2 pipeline operating pressure and potentially the size and economics of CCS projects.

The most important factor influencing the energy consumption of both CO_2 transportation in pipelines and using ships is the temperature of the cooling utility (assumed to be seawater in this study) available in the location where the CO_2 is compressed or liquefied. The results from Figure 10 and Figure 8 show that the impact of temperature is consistent for all cases and equates to around 1 % of overall energy consumption per °C across the range of temperatures and cases studied here.

Compression or liquefaction is always needed at the source of captured CO_2 emissions in CCS projects, and therefore, CCS projects located in low ambient temperature locations can be expected to benefit from lower transportation energy consumption. Figure 10 shows that this advantage can be maximized by returning the captured CO_2 directly to a storage location using compression and pipeline transportation. Interestingly, the advantage associated with a low ambient temperature location such as Melkøya in Northern Norway is also apparent when the captured CO_2 is liquefied and shipped to a storage hub located at some distance.

Allowances for the energy consumption associated with shipping (transportation fuel, re-liquefaction energy, etc.) are not calculated by the *CO2TM* and do not form part of this study. In addition, the design parameters of the pipelines and storage locations used in this study can only be taken to be indicative of the project cases they are based upon. More detailed studies would be required to make an accurate comparison of the relative performance of these different cases. However, the results presented here can provide a guide to the sensitivity of CO_2 transportation energy consumption to some important case–specific and general design parameters.

Acknowledgements

References

M. K. Chandel, L. F. Pratson, and E. Williams. Potential economies of scale in CO_2 transport through use of a trunk pipeline. *Energy Conversion and Management, 51*, 2825-2834, 2010. doi:10.1016/j.enconman.2010.06.020

M. Drescher, Ø. Wilhelmsen, P. Aursand, E. Aursand, G. de Koeijer, and R. Held. Heat Transfer Characteristics of a Pipeline for CO_2 Transport with Water as Surrounding Substance. *Energy Procedia, 37*, 3047-3056, 2013. doi:10.1016/j.egypro.2013.06.191

S. Jackson. Carbon Dioxide Transportation Energy Model. *DataverseNO,* V1, 2020a. doi:10.18710/SAIANK

S. Jackson. Development of a Model for the Estimation of the Energy Consumption Associated with the Transportation of CO_2 in Pipelines. *Energies, 13*(10), 2427, 2020b.

L. I. Langelandsvik. Modeling of natural gas transport and friction factor for large-scale pipelines: laboratory experiments and analysis of operational data. *Norwegian University of Science and Technology, Faculty of Engineering Science and Technology, Department of Energy and Process Engineering, Trondheim* (2008:221), 2008.

M. Mazzoccoli, B. Bosio, E. Arato, and S. Brandani. Comparison of equations-of-state with P-ρ-T experimental data of binary mixtures rich in CO_2 under the conditions of pipeline transport. *The journal of supercritical fluids*, 474-490, 2014.

Z. X. Zhang, G. X. Wang, P. Massarotto and V. Rudolph. Optimization of pipeline transport for CO_2 sequestration. *Energy Conversion and Management, 47*(6), 702-715, 2006. doi:10.1016/j.enconman.2005.06.001.

Hybrid Mechanistic + Neural Model of Laboratory Helicopter

Christopher Rackauckas[1]   Roshan Sharma, Bernt Lie[2]

[1]Massachusetts Institute of Technology, USA, `me@ChrisRackauckas.com`
[2]University of South-Eastern Norway, Norway, `{Roshan.Sharma,Bernt.Lie}@usn.no`

Abstract

There is a growing interest in data-driven models of dynamic systems developed from "Big Data". Data-driven models have some limitations, e.g., without data, there is no model. *Hybrid* models consisting of relatively simple mechanistic models in tandem with data-driven models offer a compromise: physics understanding and design/synthesis of control structure prior to building the system, and improved model as data becomes available.

Here, we studied a strategy to develop hybrid models based on a laboratory helicopter case study. We presented the system, with a model based on Lagrangian mechanics. Torques were assumed linear: actuator torque in voltages, and friction torque in angular velocities. Nominal model parameters were taken from a laboratory set-up. Experimental data from a laboratory process was presented; the model gave poor model fit with nominal parameters. Optimal model parameters were found based on a ballistic fit measure, and gave acceptable fit for pitch angle measurements, but poor fit for yaw angle measurements. Next, the assumption of linear controller torque was relaxed by adding a feed forward neural network block within the continuous time dynamic model. Upon training, this still gave imperfect, but considerably better prediction of the yaw angle. In a final, "model discovery" step, physically relevant alternatives to the neural network were explored.

The key idea of the paper was to illustrate a strategy for hybrid models. The current model still has some structural imperfections; these should be resolved before model validation becomes meaningful.

Keywords: neural differential equations, mechanistic model, hybrid data-driven and mechanistic model, helicopter model, control relevant model

1 Introduction

1.1 Background

Two main types of models for describing dynamic systems are (i) data-driven models (regression models, machine learning, empirical models) found by tuning parameters in some abstract mathematical structure, and (ii) mechanistic models (physics based models) which utilize balance laws based on the mechanisms that drive the change in the system coupled with semi-empiric phenomenological models (transport laws, reaction rates, etc.). These two model types have their own strengths and weaknesses. To-day there is an increasing focus on the possibility of developing *data-driven* dynamic models from "Big data".

Data-driven models put little emphasis on understanding the system, can easily be automated in computers, but require large amounts of data and do not extrapolate well.

Mechanistic models require good understanding of the physics of the system, impose limitations in model generalization, and are more complex to automate in computers. On the other hand, mechanistic models require less data, and extrapolate better.

Ideally, one would combine the best of both model types. This would imply to use simplified mechanistic models prior to the availability of data, train regression models to describe uncertain phenomenological relations and missing dynamics in the mechanistic model, and this way gradually improve the model description.

1.2 Previous work

Feed forward neural networks constitute a modern example of nonlinear regression models, i.e., a static mapping from an input signal/feature vector to an output signal which can be fitted to experimental data by modifying weight/bias parameters. Traditional use of neural networks for dynamic models are based on feed forward neural networks with auto-regressive/delayed outputs and moving average/delayed inputs. Alternatively, recurrent neural networks include internal feedback paths in layers. Both approaches are based on discrete time data/models.

Since feed forward neural networks describe a nonlinear mapping, suppose that both inputs and states of a system are known over some time horizon, together with the derivative of the states. Then one can treat the inputs and states as inputs to the neural network, and the derivatives of the states as outputs from the neural network. By tuning neural network parameters, this enables fitting the neural network to the vector field of the differential equation, leading to Neural Differential Equations (Farrell and Polycarpou, 2006; Chen et al., 2018). If all states and state derivatives are known, fitting a neural network to the vector field is straightforward. This may be the case in simulation studies for model reduction. However, for real systems with experimental data from a limited number of sensors, two problems are faced: (i) derivatives are not available and can at best be found by some smoothing/spline fitting, and (ii) derivatives can be approximated for sensor outputs, but not directly for states. This makes the problem more challenging.

Based on the ideas in (Chen et al., 2018), a set of packages for computer language Julia (Bezanson et al., 2017) are combined to fit neural differential equation models (Rackauckas et al., 2019, 2020) to experimental data, leading to differential equations that can be solved by standard differential equation solvers (Rackauckas and Nie, 2017). However, the packages aim higher: they allow for a very general mixing of mechanistic models and neural differential equation models in the same framework, with possibilities for the user to choose whether only parameters in the neural network model are tuned, or parameters in both the mechanistic model and the neural differential equation. The advantage of this approach is significant: it allows for the desirable possibility of extending mechanistic models with data-driven elements.

In this paper, we consider a mechanistic model of a laboratory helicopter at University of South-Eastern Norway; the model is used in a course on Model Predictive Control. A mechanistic dynamic model of the system can be developed as in (Gäfvert, 2001) using Lagrangian mechanics. Using sets of experimental data, a hybrid model consisting of the mechanistic model extended with neural network blocks is fitted using Julia package DiffEqFlux.jl. This allows for developing an improved model compared to the purely mechanistic model, and serves as a starting point for challenges related to hybrid models.

1.3 Overview of paper

The paper is organized as follows. Section 2 presents the laboratory helicopter, and a mechanistic model developed from Lagrangian mechanics and nominal model parameters. In Section 3, experimental data are presented, and it is shown that poor knowledge of initial values for the model with the chosen nominal parameters gives low prediction accuracy. Next, a model fitting measure is proposed (loss function), and parameters and initial values are adjusted to minimize the loss function. The model is still imperfect. In Section 4, the torque model used in the mechanistic model is replaced by a neural network block from the controller inputs (voltages) to the controller torques. Finally, in Section 5, some conclusions are drawn. Nominal model parameters and operating conditions are provided in Appendix A.

2 Helicopter mechanistic model

2.1 Laboratory helicopter

The laboratory helicopter used as a case study, is shown in Fig. 1.

A video file is available detailing the operation of the helicopter, and some elements which makes it complicated to achieve a perfect mechanistic model of the helicopter.[1]

2.2 Geometry of helicopter

Consider the helicopter in Fig. 2, with upward-pointing z axis. *Origo* of the body-fixed coordinate system is in the

[1]https://web01.usn.no/~roshans/mpc/videos/Heli_deadband_effect.mp4

longitudinal inertial axis of the helicopter. The laboratory helicopter is hinged to the ground; in Fig. 2, the pivot point is located in the body-fixed origo.

If the helicopter body is elevated a distance h above the pivot point, differential mass $dm(r)$ at position r along the longitudinal axis of the helicopter is given by coordinates z_r and ξ_r,

$$z_r = r\sin\theta + h\cos\theta$$
$$\xi_r = r\cos\theta - h\sin\theta.$$

The angle of nose raise θ is termed the *pitch* angle. Seen from a birds-eye view, the helicopter can also rotate in the plane by the *yaw* angle ψ, Fig. 3.

The positions in the plane are

$$x_r = \xi_r \cos\psi$$
$$y_r = \xi_r \sin\psi.$$

In summary, the position of mass $dm(r)$ is given by Cartesian coordinates (x_r, y_r, z_r)

$$x_r = (r\cos\theta - h\sin\theta)\cos\psi$$
$$y_r = (r\cos\theta - h\sin\theta)\sin\psi$$
$$z_r = r\sin\theta + h\cos\theta,$$

where the generalized coordinates are $\mathsf{x} = (\theta, \psi)$.

2.3 Kinetic energy of helicopter

With linear velocities $v_x \triangleq \frac{dx_r}{dt}$, $v_y \triangleq \frac{dy_r}{dt}$, and $v_z = \frac{dv_z}{dt}$, and introducing angular velocities $\omega_\theta \triangleq \frac{d\theta}{dt}$ and $\omega_\psi \triangleq \frac{d\psi}{dt}$, the squared velocity $v_r^2 = v_x^2 + v_y^2 + v_z^2$ can be expressed by the generalized velocities $\mathsf{v} = (\omega_\theta, \omega_\psi)$ as

$$v_r^2 = r^2\left(\omega_\theta^2 + \cos^2\theta \cdot \omega_\psi^2\right) + h^2\left(\omega_\theta^2 + \sin^2\theta \cdot \omega_\psi^2\right)$$
$$- 2rh\sin\theta\cos\theta \cdot \omega_\psi^2.$$

Combining the rotation of the helicopter body around the pivot point with body center moment of inertia J_{bc} and mass center distance r_c, and rotation around the shaft with moment of inertia J_s gives a total kinetic energy expressed as

$$K(\mathsf{x}, \mathsf{v}) = \frac{1}{2}\mathsf{v}^T\mathsf{M}(\mathsf{x})\mathsf{v}$$

where the mass matrix $\mathsf{M}(\mathsf{x})$ can be shown to be diagonal with

$$\mathsf{M}_{11}(\mathsf{x}) = J_{bc} + mr_c^2 + mh^2$$
$$\mathsf{M}_{22}(\mathsf{x}) = J_{bc}\cos^2\theta + m(r_c\cos\theta - h\sin\theta)^2 + J_s.$$

2.4 Potential energy of helicopter

Setting the potential energy to zero in the pivot point, the potential energy of the helicopter becomes

$$P = \int_m gz_r dm(r)$$
$$\Downarrow$$
$$P = g\int_m (r\sin\theta + h\cos\theta)dm(r),$$
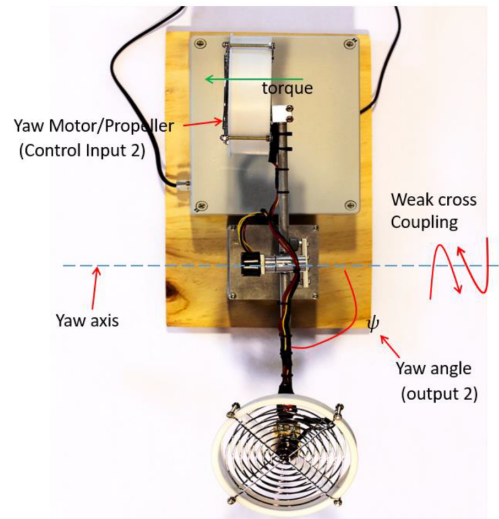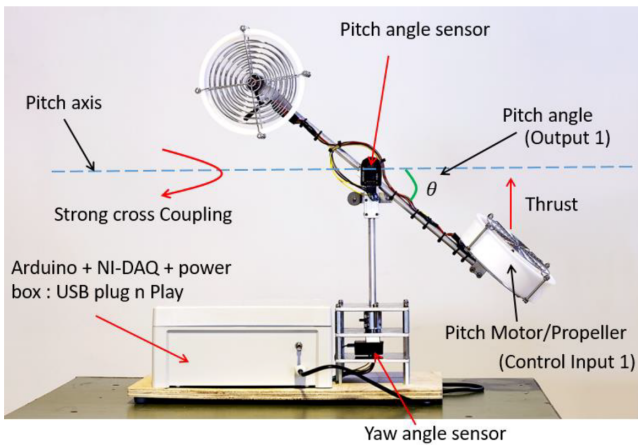
Figure 1. Laboratory helicopter, (Sharma, 2020).
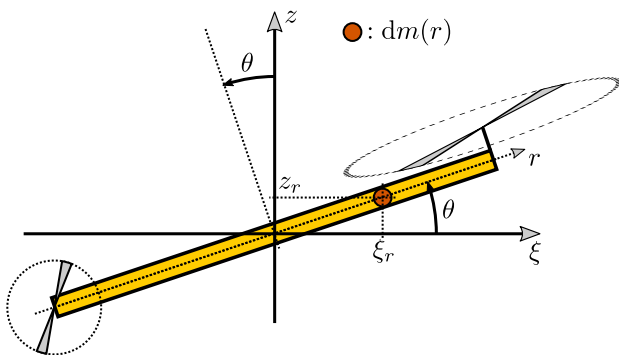
Figure 2. Helicopter in side profile, pitch angle θ.
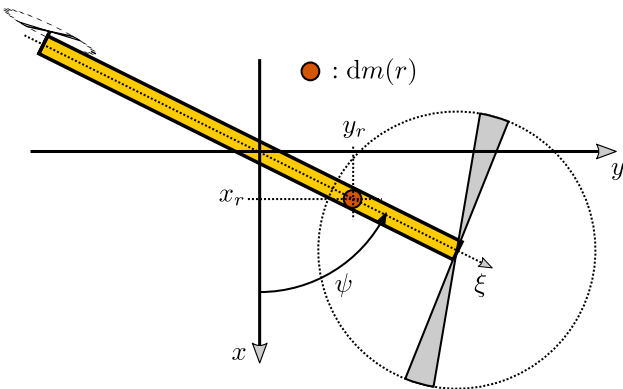
Figure 3. Helicopter in birds-eye view, yaw angle ψ.

which can be expressed as

$$P(\mathsf{x}) = mg\left(r_c \sin\theta + h\cos\theta\right).$$

2.5 Helicopter torques

We need the generalized force $\mathsf{F} = \left(T_\theta, T_\psi\right)$, which is given by some phenomenological relation

$$\mathsf{F} = \mathsf{F}\left(\mathsf{x}, \mathsf{v}, u\right).$$

Here, $u = \left(u_\theta, u_\psi\right)$ contains the voltages u_θ and u_ψ applied to the rotors. In mechanistic models, a proposal for F could be

$$\mathsf{F} = Ku - D\mathsf{v}$$

with a full motor gain matrix K (Sharma, 2020)

$$K = \left(\begin{array}{cc} K_{\theta,\theta} & -K_{\theta,\psi} \\ K_{\psi,\theta} & -K_{\psi,\psi} \end{array}\right),$$

and a diagonal damping friction matrix D

$$D = \left(\begin{array}{cc} D_\theta & 0 \\ 0 & D_\psi \end{array}\right).$$

In practice, this linear description may be too simple: the system may exhibit dead-band thus invalidating the term Ku. Presence of nonlinear friction such as stiction or quadratic terms in v invalidate the term $D\mathsf{v}$. Torsional effects would imply missing potential energy in x. The system may even exhibit backlash, and thus make it necessary to introduce extra states z with generalized force $\mathsf{F}(\mathsf{x}, \mathsf{v}, z, u)$ (Lichtsinder and Gutman, 2016, 2019).

Particular problems related to how the laboratory helicopter is constructed, are discussed in link `https://web01.usn.no/~roshans/mpc/videos /Heli_deadband_effect.mp4`: power from u is transmitted to the helicopter via wires in the shaft, with twisting of these wires. This design may give rise to some dead-band, but it may perhaps also give rise to some resisting torsion spring torque.

2.6 DAE formulation of model

We now introduce the *Lagrangian* L defined as

$$\mathsf{L}(\mathsf{x},\mathsf{v}) \triangleq K(\mathsf{x},\mathsf{v}) - P(\mathsf{x}).$$

Defining the generalized momentum $\mathsf{p} \triangleq \partial \mathsf{L}/\partial \mathsf{v}$, a DAE formulation of the Euler-Lagrange equation can be posed as

$$\frac{d\mathsf{x}}{dt} = \mathsf{v}$$
$$\frac{d\mathsf{p}}{dt} = \frac{\partial \mathsf{L}}{\partial \mathsf{x}} + \mathsf{F}$$
$$\mathsf{p} = \mathsf{M}(\mathsf{x})\mathsf{v}.$$

Here, the momentum p contains angular momenta, $\mathsf{p} = (\mathsf{p}_\theta, \mathsf{p}_\psi)$, while term $\frac{\partial \mathsf{L}}{\partial \mathsf{x}} = \left(\frac{\partial \mathsf{L}}{\partial \theta}, \frac{\partial \mathsf{L}}{\partial \psi}\right)$.

2.7 ODE formulation of model

If we eliminate the momentum, we can rewrite the model as implicit ODEs,

$$\frac{d\mathsf{x}}{dt} = \mathsf{v}$$
$$\mathsf{M}(\mathsf{x})\frac{d\mathsf{v}}{dt} = -\frac{\partial \mathsf{p}}{\partial \mathsf{x}}\mathsf{v} + \frac{\partial \mathsf{L}}{\partial \mathsf{x}} + \mathsf{F}.$$

The ODE form of the model can then be expressed as

$$\mathsf{M}_{11}\frac{d\omega_\theta}{dt} = -\omega_\psi^2 \left(\left(J_{bc} + m\left(r_c^2 - h^2\right)\right)\cos\theta\sin\theta\right.$$
$$\left. - r_c h\left(\cos^2\theta - \sin^2\theta\right)\right)$$
$$- mg\left(r_c\cos\theta - h\sin\theta\right) + T_\theta$$

$$\mathsf{M}_{22}\frac{d\omega_\psi}{dt} = 2\omega_\psi\omega_\theta\left(\left(J_{bc} + m\left(r_c^2 - h^2\right)\right)\cos\theta\sin\theta\right.$$
$$\left. + mr_c h\left(\cos^2\theta - \sin^2\theta\right)\right) + T_\psi$$

with the trivial additional equations

$$\frac{d\theta}{dt} = \omega_\theta$$
$$\frac{d\psi}{dt} = \omega_\psi.$$

In (Sharma, 2020), a simplified model is used with $h \equiv 0$, neglecting the shaft moment of inertia J_s, etc.

3 Preliminary model fitting

3.1 Experimental data

Voltages for pitch motor u_θ and yaw motor u_ψ used in one particular experiment[2] with $N_d = 13\,100$ datapoints sampled at $\Delta t = 0.01$ are displayed in Fig. 4; notice the different ordinate axes on each side of the plot.

Observed pitch angle θ and yaw angle ψ from the particular experiment are shown in Fig. 5. Again, notice the use of different ordinate axes on each side of the plot.

Pitch angle θ and yaw angle ψ from experiments with nominal model parameters are displayed in Fig. 6.

---

[2]https://web01.usn.no/~roshans/mpc/videos/Heli_deadband_effect.mp4
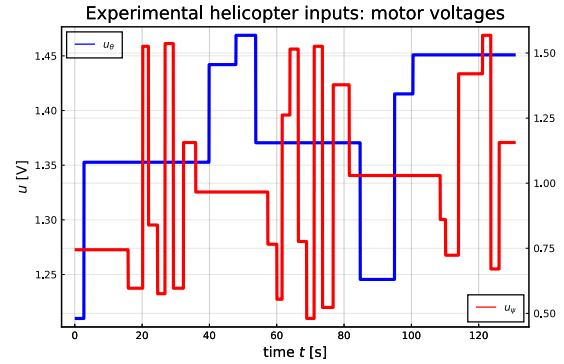
Figure 4. Pitch motor voltage u_θ and yaw motor u_ψ experiments.
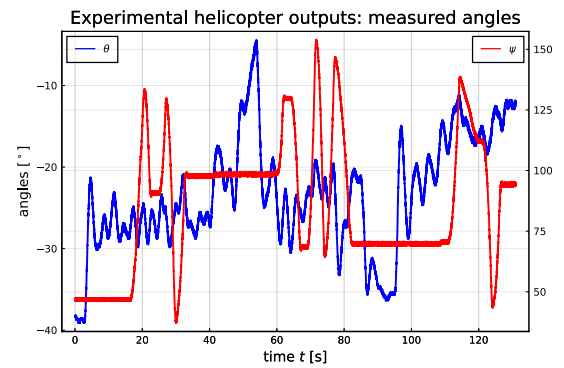
Figure 5. Pitch angle θ and yaw angle ψ from experiments.
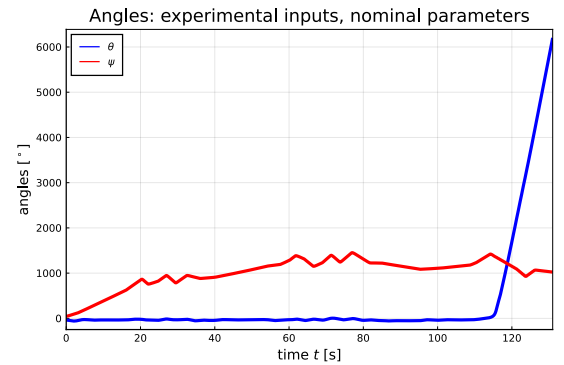
Figure 6. Model pitch angle θ and yaw angle ψ based on nominal model parameters, using experimental inputs.

Table 1. Fitted parameters for laboratory helicopter.

Parameter	Nominal value p_0	Fitted value p^*
r_{c}	1.5 cm	2.2 cm
h	0 cm	0.1 cm
m	0.5 kg	0.72 kg
J_{bc}	0.015 kg m^2	0.01 kg m^2
J_{s}	0.005 kg m^2	0.0017 kg m^2
$K_{\theta,\theta}$	0.055 Nm/V	0.105 Nm/V
$K_{\theta,\psi}$	0.005 Nm/V	0.0017 Nm/V
$K_{\psi,\theta}$	0.15 Nm/V	0.03 Nm/V
$K_{\psi,\psi}$	0.20 Nm/V	0.043 Nm/V
D_θ	0.01 Nm/(rad/s)	0.014 Nm/(rad/s)
D_ψ	0.08 Nm/(rad/s)	0.11 Nm/(rad/s)
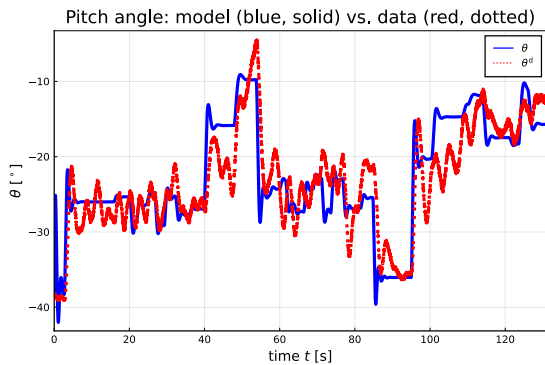
Figure 7. Pitch angle θ: comparing model angle with optimized parameters (blue, solid) and measured experimental angle.

3.2 Preliminary model fitting

We now tune model parameters p to minimize the *loss* function ℓ

$$
\begin{aligned}
\ell(p) &= \sqrt{\frac{1}{n_y \cdot N_{\mathrm{d}}} \sum_{k=1}^{N_{\mathrm{d}}} \left[\left(\theta_k - \theta_k^{\mathrm{d}}\right)^2 + \left(\psi_k - \psi_k^{\mathrm{d}}\right)^2 \right]} \\
&= \frac{\|y - y^{\mathrm{d}}\|_2}{\sqrt{n_y \cdot N_{\mathrm{d}}}} = \sigma_\varepsilon
\end{aligned}
\tag{1}
$$

where θ_k is the model pitch angle and ψ_k is the model yaw angle, while θ_k^{d} and ψ_k^{d} are logged data of the same angles, and $n_y = \dim y_k = 2$. It follows that $\ell(p)$ is the average standard deviation σ_ε in each measured angle. The model angles vary with the model parameters p; $\theta_k = \theta_k(p)$ and $\psi_k = \psi_k(p)$. In Eq. 1, θ_k and ψ_k are found from *ballistic* simulation/single shooting of the model from the initial conditions, and not from multiple shooting as in prediction error methods.

In the model fitting, the 4 initial states and 11 model parameters are tuned to minimize $\ell(p)$ using method `bboptim` with default settings in package `BlackBoxOptim.jl` for Julia. Table 1 shows the fitted values of model parameters.

Pitch angle θ from the model and from experimental data with optimized model parameters are given in Fig. 7.
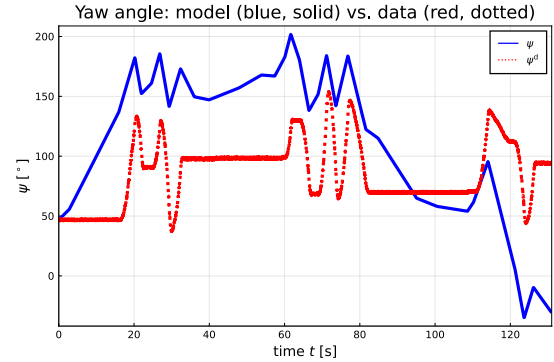
Figure 8. Yaw angle ψ: comparing model angle with optimized parameters (blue, solid) and measured experimental angle.

Yaw angle ψ from model and from experimental data with optimized model parameters are displayed in Fig. 8.

Because a mechanistic model with given parameter set is used, and because poor model fit is achieved for the yaw angle, it is not relevant to consider model validation: the model structure is too limited to allow for good fit. The optimal loss function is $\ell(p^*) \approx 92$, with the error mainly in the yaw angle.

4 Hybrid model

4.1 Neural torque extensions

In Section 2.5, we considered a helicopter torque given as $\mathsf{F} = Ku - D\mathsf{v}$. We now modify this torque expression to

$$
\mathsf{F} = Ku - D\mathsf{v} + \mathrm{FNN}(u; p)
\tag{2}
$$

where we keep the optimal parameter values in Table 1, but re-fit the model by tuning parameters p in $\mathrm{FNN}(\cdot)$ consisting of weights and biases in a two layer Feed forward Neural Net with 2 inputs, 16 nodes in the hidden layer and a $\tanh^{-1}(\cdot)$ activation function, and 2 outputs in the linear output layer. The tuning of the neural net is carried out in two phases: in the first phase, gradient search is done using an ADAM method of Julia package DiffEqFlux.jl, using the global loss function in Eq. 1 of Section 3.2. In a second, polishing step, we use a quasi-Newton method BFGS method to fine tune the parameters of the neural net.

After fitting the neural network, the pitch angles are as in Fig. 9 — which should be compared to Fig. 7.

The model fitting of the yaw angle is as in Fig. 10 — which should be compared to Fig. 8.

The resulting loss function after fitting the 16 node FNN is $\ell(p^*) \approx 9.2$ after phase 2 with the quasi-Newton BFGS method ($\ell(p^*) \approx 55.4$ after the ADAMS gradient search of phase 1). Using a 32 node FNN leads to $\ell(p^*) \approx 9.3$ after phase 2 ($\ell(p^*) \approx 35$ after phase 1); thus going to 32 nodes does not improve the fit.

We may also modify the torque expression to include a neural network which both depends on input voltage u and
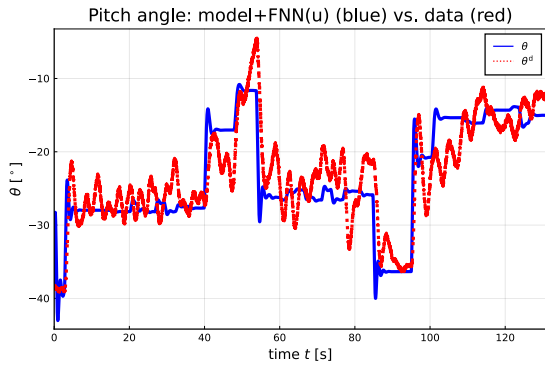
Figure 9. Pitch angle with torque modeled as in Eq. 2 with additive term $\text{FNN}(u; p)$.

Figure 10. Yaw angle with torque modeled as in Eq. 2 with additive term $\text{FNN}(u; p)$..

states x, v:

$$\mathsf{F} = K u - D \mathsf{v} + \text{FNN}(u, \mathsf{x}, \mathsf{v}; p). \tag{3}$$

Now we have a neural network with 6 inputs (u, x, v) and two outputs x, with a hidden layer with 16 nodes. The resulting loss function is $\ell(p^*) \approx 11.1$ after phase 2 ($\ell(p^*) \approx 14.1$ after phase 1). With the limited available data, this extension doesn't improve the model fit: the pitch angle prediction is somewhat poorer, while the yaw angle prediction is slightly better.

4.2 Equation discovery

When finding the model augmented with a neural network addition to the torque F, we know the experimental input u and compute the resulting solution x. We can then post process the solution and compute $\text{FNN}(u; p)$. This gives us the time series of u and $\text{FNN}(u; p)$, which we will use for "equation discovery". The idea is simply to propose a regression model between u and $\text{FNN}(u; p)$, and then fit this regression model. In such a regression model, we will postulate a large number of basis function which may have a physical origin. As an *example*, we could postulate a regression model of form

$$\text{FNN}(u; p) = c_1 + c_2 u_\theta + c_3 u_\psi + c_4 u_\theta^2 + c_5 u_\psi^2 + c_6 u_\theta u_\psi \cdots \tag{4}$$

Figure 11. Pitch angle with regression approximation of $\text{FNN}(u; p)$.

In this case, we have a linear regression model that can easily be fitted to the data set $(u, \text{FNN}(u; p))$. Some of the data set can be used for training the model, and some of it can be used for validation. This way, it is possible to "discover" which of the basis functions/equations that are relevant to describe the fitted neural network part of the torque.

To automate this process, package `DataDrivenDiffEq.jl` is used. A set of basis functions are proposed in line with the regression equation in Eq. 4 using the `Basis()` constructor of package `DataDrivenDiffEq.jl`. Next, the optimizer `SINDy` (Sparse Identification of Nonlinear Dynamics) of this package is used to "discover" the relevant equations that explain the neural network. If we denote by $\text{FNN}(\cdot)$ the output data from the neural network, by $\Phi(u_\theta, u_\psi)$ denote the regressor data (i.e., the RHS terms in Eq. 4), and by $p = (c_1, c_2, \ldots)$ denote the vector of unknown parameters, we can pose a multi-objective problem combined by weight λ as

$$p^* = \arg\min_p \left(\left\| \text{FNN}(\cdot) - \Phi(u_\theta, u_\psi) p \right\|_2 + \lambda \left\| p \right\|_1 \right). \tag{5}$$

Here, the use of 1-norm in the regularization term, $\left\| p \right\|_1$, tends to favor sparse solutions, and the weight λ is varied to give a number of alternative models. The model choice is then based on the Akaike Information Criterion. Finally, this set of discovered equations can replace the neural network in the expression for F.

We find that

$$\text{FNN}(u; p)_1 \approx -4.37 \cdot 10^{-4} \cos(u_\psi) + 4.02 \cdot 10^{-4} \sin(u_\psi)$$
$$\text{FNN}(u; p)_2 \approx -1.35 \cdot 10^{-2} \cos(u_\psi) + 7.74 \cdot 10^{-3} u_\psi^2.$$

By replacing the neural network with the approximate, "discovered" equations, the model fit is as in Figs. 11 and 12.

When comparing Figs. 9 and 10 to Figs. 11 and 12, observe that the pitch approximation is almost identical, while the yaw approximation is good — but a little bit off that from $\text{FNN}(u; p)$.

Figure 12. Yaw angle with regression approximation of $FNN(u; p)$.

Table 2. Summary of loss function values. For Mechanistic model of F, a black box optimizer is used. For hybrid models of F, a gradient based phase 1 is used ($\ell(p^*)_1$) followed by a quasi-Newton phase ($\ell(p^*)_2$). The index on the neural networks indicate the number of nodes, 16 or 32.

Case	$\ell(p^*)_1$	$\ell(p^*)_2$
Mechanistic		92
Hybrid, $FNN(u; p)_{16}$	55.4	9.2
Hybrid, $FNN(u; p)_{32}$	35	9.3
Hybrid, $FNN(u, x, v; p)_{16}$	14.1	11.1

5 Conclusions and Future work

The key problem of fitting mechanistic models to experimental data has been considered, based on a simple model of a laboratory helicopter and experimental data from a lab rig. As in all engineering practice, the model structure of the original mechanistic model limits how well it is possible to fit the model to real data. Still, a mechanistic model gives important information about the system and may often be used for control design with some success, even when a perfect model fit is not possible. The reason why control design works for imperfect models is that feedback to some degree corrects for the effect of model error.

It is of interest to *improve* on a mechanistic model as more data becomes available. Thus, we do not want to replace the mechanistic model by a 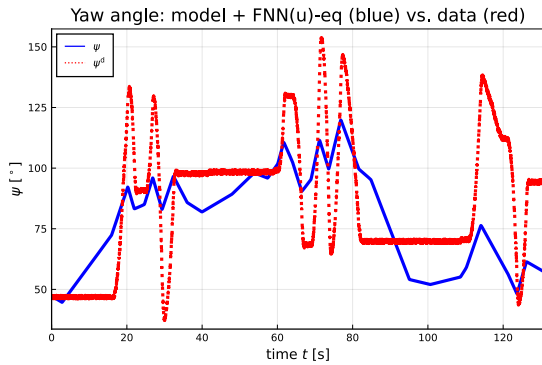completely data driven model, but instead augment the mechanistic model with data driven elements as more information becomes available. An obvious way to do this is to add regression type blocks in the mechanistic model, both in phenomenological laws and possibly also in adding more states to the system with regression type vector fields. One possible data driven model structure is a neural network. The challenge with including regression blocks in dynamic models is how to integrate these in an efficient way so that differential equation solvers can handle them, so that it is possible to take advantage of GPUs, and how to make accessible model gradients for training the network within the dynamic model.

In this paper, we have used packages within the DifferentialEquations.jl package and the Flux.jl package ecosystems of Julia to achieve this. The original, fitted mechanistic model is good for pitch angle predictions, but poor for yaw angle predictions. By adding a torque $FNN(u; p)$ computed via a trained neural network, the yaw angle prediction improves considerably. Specifically, the loss function used for model fit is reduced from $\ell(p^*) = 92$ with the original model to $\ell(p^*) = 9.2$ when the neural network torque term with 16 internal nodes has been added. In our case of model and available data, it appears that there is no gain in increasing the number of internal nodes, or to

include mapping $FNN(u, x, v; p)$ from the states; this does not reduce the loss function value. The loss function values are summarized in Table 2.

With multiple candidate models, it is common to use model validation on independent data to choose model. The key reason for doing this is to avoid model overfitting/fitting noise. With the data we have available and the tested neural network mappings, the model fit is still relatively poor, so there is little sense in doing model validation at the current stage.

For practical implementation, a neural network is not the most convenient model, as it is rather complex with many parameters. Instead, we would like to replace the neural network by a regression approximation using more physically relevant basis functions. This is what the *Equation discovery* part of the Julia eco-system allows for. As seen, such equation discovery can lead to dramatically reduced model complexity compared to a neural network.

Overall, the tools offered in Julia give a first glimpse into what future modeling practice will look like: one starts by developing an efficient, and relatively simple mechanistic model for design/control synthesis/... As more data become available, the model can be improved by adding unstructured data driven terms (e.g., neural networks) which subsequently are given a physical interpretation by equation discovery.

Future work with the helicopter model will include experiments with more data, neural networks with added states, equation discovery with more extensive sets of basis functions, and testing how the improved model influences the controller performance.

A Nominal parameters and operating conditions

Nominal parameters for the mechanistic model are given in Table 3.

Nominal operating conditions for simulations are chosen as in Table 4.

References

Jeff Bezanson, Alan Edelman, Stefan Karpinski, and Viral B. Sha. Julia: A Fresh Approach to Numerical Computing. *SIAM Review*, 49(1):65–98, 2017. doi:10.1137/141000671.

Table 3. Parameters for laboratory helicopter.

Parameter	Value	Description
r_c	1.5 cm	Distance between pivot point and center of mass.
h	0 cm	Elevation of helicopter from pivot point.
m	0.5 kg	Mass of helicopter
J_{bc}	0.015 kg m^2	Moment of Inertia of body about center of mass.
J_s	0.005 kg m^2	Moment of Inertia of shaft.
g	9.81 m/s^2	Acceleration of gravity.
$K_{\theta,\theta}$	0.055 Nm/V	Torque constant on pitch coordinate from pitch motor.
$K_{\theta,\psi}$	0.005 Nm/V	Torque constant on pitch coordinate from yaw motor.
$K_{\psi,\theta}$	0.15 Nm/V	Torque constant on yaw coordinate from pitch motor.
$K_{\psi,\psi}$	0.20 Nm/V	Torque constant on yaw coordinate from yaw motor.
D_θ	0.01 Nm/(rad/s)	Friction torque damping coefficient in pitch coordinate.
D_ψ	0.08 Nm/(rad/s)	Friction torque damping coefficient in yaw coordinate.

Table 4. Initial operating conditions for laboratory helicopter.

Quantity	Value	Description
$\theta(0)$	$\frac{\pi}{4}$ rad	Initial pitch angle.
$\psi(0)$	0 rad	Initial yaw angle.
$\omega_\theta(0)$	-0.1 rad/s	Initial pitch angular velocity.
$\omega_\psi(0)$	0.1 rad/s	Initial yaw angular velocity.
$u_\theta(t)$	0 V	Pitch motor voltage.
$u_\psi(t)$	0 V	Yaw motor voltage.

Christopher Rackauckas, Yingbo Ma, Julius Martensen, Collin Warner, Kirill Zubov, Rohit Supekar, Dominic Skinner, and Ali Ramadhan. Universal differential equations for scientific machine learning. *arXiv CoRR*, abs/2001.04385, 2020. https://arxiv.org/abs/2001.04385.

Roshan Sharma. Two degrees of freedom (2-dof) helicopter. Laboratory problem in course iia 4117 model predictive control, University of South-Eastern Norway, 2020.

Tian Qi Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural ordinary differential equations. *arXiv CoRR*, abs/1806.07366, 2018. http://arxiv.org/abs/1806.07366.

Jay A. Farrell and Marios M. Polycarpou. *Adaptive Approximation Based Control. Unifying Neural, Fuzzy and Traditional Adaptive Approximation Approaches*. Wiley-Interscience, Hoboken, New Jersey, 2006.

G.M. Gäfvert. Modelling of the eth helicopter laboratory process. Technical Report TFRT-7596, Department of Automatic Control, Lund Institute of Technology (LTH), 2001.

Arkady Lichtsinder and Per-Olof Gutman. Closed-form sinusoidal-input describing functionfor the exact backlash model. *IFAC-PapersOnLine*, 49(18):422–427, 2016. doi:https://doi.org/10.1016/j.ifacol.2016.10.202.

Arkady Lichtsinder and Per-Olof Gutman. On the dual properties of friction andbacklash in servo control systems. *IFAC-PapersOnLine*, 52(16):340–345, 2019. doi:https://doi.org/10.1016/j.ifacol.2019.11.803.

Christopher Rackauckas and Qing Nie. DifferentialEquations.jl — A Performant and Feature-Rich Ecosystem for Solving Differential Equations in Julia. *Journal of Open Research Software*, 5(15), 2017. doi:10.5334/jors.151.

Christopher Rackauckas, Mike Innes, Yingbo Ma, Jesse Bettencourt, Lyndon White, and Vaibhav Dixit. Diffeqflux.jl - A julia library for neural differential equations. *arXiv CoRR*, abs/1902.02376, 2019. http://arxiv.org/abs/1902.02376.

Decision tree for enhancing maintenance activities with drones in the mining business

Madeleine Martinsen [1], Erik Dahlquist [2], Anders Lönnermark [3], Örjan Säker [4]

[1]ABB Industrial Automation, Industrial PhD student at MDH, Västerås Sweden,
madeleine.martinsen@mdh.se
[2]MDH, Mälardalens University, Västerås Sweden, erik.dahlquist@mdh.se
[3]RISE, Research Institutes of Sweden, Borås Sweden, anders.lonnermark@ri.se
[4]Epiroc, Rocktec Automation Rock Drills AB, Örebro Sweden, orjan.saker@epiroc.com

Figure 1. The open pit Aitik, Boliden.

Abstract

A drone (UAV, unmanned aerial vehicle) is no longer a toy, it is gaining bigger and bigger terrain in the industry as an everyday working tool. Equipped with sensors, thermal cameras, components and system software it most likely will be part of the solutions that continues streamlining the mining operations in the future. Data gathering of signals from sensors mounted on drones and other mining equipment such as mining vehicles creates conditions for monitoring, analysis and warning of possible risks and suggests how these can be avoided in due time. The experimental drone study conducted at an open pit mine Aitik, Boliden (Figure 1) in Sweden will be presented in this paper. Aitik is today the world's most efficient copper (Cu) open pit mine. The authors propose decision trees to support and enable the transformation into a completely autonomous mining operation. In combination with deep learning (DL), pattern recognition and artificial intelligence (AI) applications, creates the puzzle pieces to support mining operations to further increase their productivity and safety.

Keywords: maintenance, inspections, drones, mining operations, decision tree, artificial neuron

1 Introduction

1.1 Background

Metals and minerals are part of our daily lives without them we would not be able to for example use mobiles, brew a cup of coffee or heat dinner in the oven. This makes us dependent on the world's mines. The mining operation consists of several subprocesses like drilling, charging, blasting, loading, crushing, transportation of ore, grinding, flotation, thickening, filtering, drier, smelter, converter as shown in Figure 2. A time-consuming, expensive and sometime dangerous activity is to conduct maintenance of the entire mining processes consisting of multiple equipment's requiring continuous inspections. Performing maintenance on for example the crusher and the grinding machine demands they are stopped for manually inspections. A job that can take up to a week to perform and is repeated a number of times a year. Such stop entails a cost of roughly 40 00 USD per hour when production is at a standstill. In any case, maintenance is an important part to ensure a safe mining operation and maintain a high productivity. In addition, the operation experiences dangerous situation like for example fires resulting in longer production stops. A fire in a mine may cause both deaths and major issues with respect to production that can be very costly.

Roughly one fire incident occurs per week in the Swedish mines (Hansen, 2011, 2015) with similar frequencies also internationally (Hansen, 2018). 80 % of the fires are caused by mining vehicles (Martinsen et al., 2019). By detecting faulty equipment's such as overheated cables, motors, brakes, pressure changes on tires, oil leakage in engines or hydraulics or similar faults, escalating of fires can be avoided. A further challenge arises when mining production is planned to increase significantly while impairment in productivity or safety is not an option. This is the case for Aitik where production is planned to increase from 36-million-ton ore per year to 45 million ton.

In view of the above stated challenges, the research project aims to evaluate the possibility of introducing drones as a working tool to streamline several of the mining maintenance inspection activities and, if relevant, replace daily duties that are being conducted manually. In addition, the study aims to investigate the possibility of defining decision trees to support fault and root cause analysis from the gathered sensor signals. With the goal to warn, in good time, concerned personnel. Images taken on objects during the drone route and data gathered from sensors attached on the drone and mining equipment's are compared and in the event of deviation from normal operation values, the operators could be warned betimes.

1.2 Drone application area in the mining industry

As Figures 3 shows, the market and applications for drones has exploded (Drone Insight Industry, The Drone Market Report 2019-2024). A number of benefits, where companies have introduced drones have been able to demonstrate several advantages such as:
- Time savings
- Enhancement of the productivity
- Supporting the digitization strategy
- Possibility to increase revenue
- Increasing the control and supervision function
- Minimizing manually duties
- Contribute to a safer workplace

This creates curiosity among mining companies to use drones equipped with e.g. IR cameras, gas or other relevant sensors to gather data from these and support the subprocesses like:

- Inspection of power distributions equipment
- Inspection of shafts & vessels
- Inspection of mine hoist ropes
- Inspection of infrastructure (e.g. roads & ramps)
- Inspection of equipment e.g. conveyer belts
- Gas measurement after blasting
- Pit and dump management
- Communication and checking daily/weekly plans
- Haul route surface optimizations
- Haul road, dump and pit design
- Up-to-date surfaces for optimized blast designs
- Pre- and post-blast data
- Stockpile management, 3D mapping
- Grade control & exploration planning
- Drainage and water management
- Watershed, drainage basin, and water flow mapping
- Thermal detection of ground water inflows
- Tailing dams management
- Construction feasibility studies
- Leach pad, dam wall, and platform construction quality control
- Construction progress monitoring & reporting
- Geophysical & watershed/catchment area modeling
- Land usage reporting
- Heritage and environmental management
- Erosion detection, Inspection *of* rock mechanic
- Vegetation change tracking
- Security, corridor, and boundary surveillance
- Trolley equipment for the electrified truck route
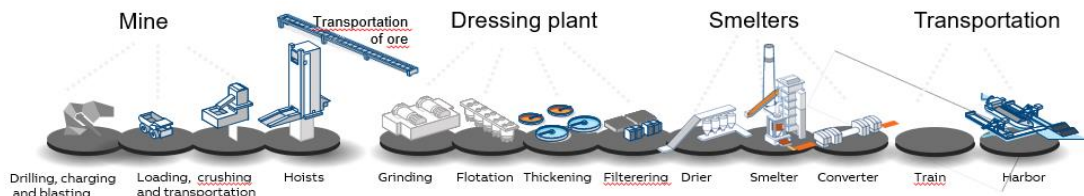
Figure 2. The Mining Process, picture by ABB.

Figure 3. The Drone Market 2018, Manufactures, Softwares, Systems, Suppliers, Medias, Insurance.

2 Methodology

2.1 Material and methods

A sequence of tests has been carried out to investigate the potential and conditions with drones to replace or support some of the current preventive maintenance inspections and manual duties in the open pit. For the tests the following mining activities and areas were chosen: gas measurement after blasting, inspection of press filter area by flying in the concentrator silo and water pipes, inspection in flotation, primary mill section, crusher and the trolley line (electrified). Several of these are hard-to-reach areas that pose a danger and risk for an employee to inspect. In addition, maintenance personnel are transported with diesel-powered mining vehicles, a trip that can take up to 30 minutes to just travel from one side to the other of the open pit. During the inspections, the test; with the drones, aimed at investigating for example:

- Discolorations that could be a hint of rust on e.g. pipes in hard to reach areas in e.g. the flotation building
- Damage to the equipment and thus verify the need for spare part change
- Bolt movements

- Flying inspection with thermal camera of moving trucks to detect e.g. oil leakage or elevated engine temperatures, worn tires

A few drones were selected for the test, see Figure 4 - 7, and the features defined in Table 1, were evaluated when selecting proper drones. The selected drones had different or similar capacities to meet those demands.

Table 1. Drone requirements.

Requirements	Comment
Suitable for an open pit operation	It is a dirty and hazardous environment.
Battery solutions	Some activities demand that the drone can operate for a longer sequence.
Temperature resistant	The weather can reach minus 30 °C or even colder and snow.
Payload capacity	Different payload will be attached for the different inspection activities like for example thermal camera and gas sensor.
Avoidance solutions	Confined spaces will be inspected

Figure 4. Flyability Elios 2.

Figure 5. DJI Matrice 210 XT2.

Figure 6. DJI Matrice 210 Z30.

Figure 7. DJI Mavic 2 Zoom.

The Flyability Elios, Figure 4, was selected for inspecting places and objects where few or no inspections have been carried out in the past since some areas are difficult, or even impossible, to reach by a human or not secure enough. The advantage of this drone is that it is less sensitive to impact on the inspected areas, as the drone is encased in a steel structure.

2.2 Theory

This document aims to propose possible solutions that improve complex and time-consuming inspection processes with the goal to prevent dangerous situations, such as fires, from occurring. For achieving accurate and earlier discoveries of equipment deficiencies we introduce Bayesian Networks, BN, (Berthold et al., 1999), decision trees, for analysis of the trend (Craven, 1997) development of features and decision support.

Neural networks (McCulloch et al., 1943; Haykin, 1999) represent a brain metaphor that process information. The biological neuron ideates the

elementary unit of any biological nervous system (McCulloch et al. 1943; Amit et al., 2003; Berthold et al., 1999; Zurada, 1992). They appear to be organized in structures where, after sufficient learning period, collaborate to solve a high number of complex tasks. Further neural computing refers to a pattern recognition (Tsai, 2009) methodology for machine learning (ML) where an artificial neural network (ANN) in Figure 8 is the resulting model from neural computing (McCulloch et al., 1943). Neural networks are programmed to:

- Store
- Recognize
- Associatively retrieve patterns or database entries
- To solve combinatorial optimizations problem
- To filter noise from measurement data
- To control ill-defined problems
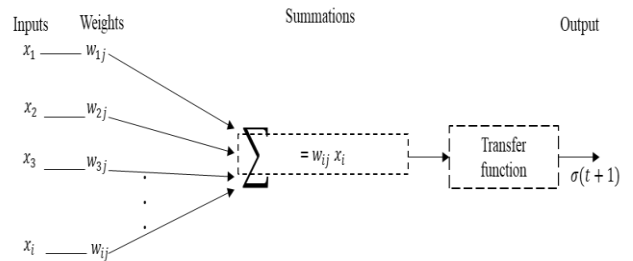
Figure 8 The artificial neuron.

Neural networks have gained eminent interest due to its forecasting applications, clustering and business classifications applications for prediction and pattern recognition in connection to their ability to "learn" from the data, their nonparametric nature (i.e., no rigid assumptions), and their capacity to generalize.

$$\sigma(t+1) = f(\sum_{i=1}^{n} w_i x_i(t)) \qquad (1)$$

By training, the neural network method gradually calculates, equation (1) (including repeated iteration or cumulative calculation) the weights, w_{ij} , the neural network connected. In our study, the inputs, x_{ij} , are represented by the values from the signal gathered from the mining equipment's different sensors like. temperature and gases (hydrocarbons, NOx…) being attached e.g. to a mining vehicle but also on a drone where the temperature can be measured by the FLIR camera. Then combining the gathered input signals root cause analyses (RCA) can be carried out supported by a graphical structure of BN decision tree model as shown in Figure 9.
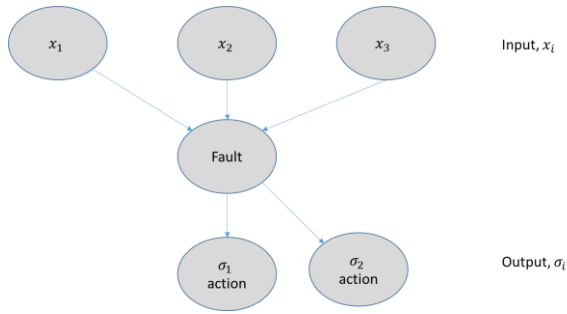
Figure 9. Decision Tree.

Neural networks have proved to be efficient to use for estimating sampled functions when we do not know the form of the function. The abilities, pattern recognition and function estimation constitute ANNs. It is a general methodology used for many functions, e.g. autonomous driving (Kasper et al., 2012) and process control (G. Weidl et al., 2005 and 2008, Abraham et al., pp. 221-230, 2002) for condition monitoring, RCA and decision support of complex processes.

The gas tests carried out (Martinsen et al., 2019) showed possibilities to detect oil leakage, oil mist and risks for fires or toxic substances on mining vehicles before they cause harm. Different sensors detected hydrocarbons, NOx, high temperatures at "risk positions" both at vehicles and stationary. Some 20 different types of sensors were evaluated. To detect gases (Martinsen et al., 2020) at an early stage with adequate concentration the study showed the need to suck gas from critical positions, e.g. in the engine, close to hydraulics in the drilling machine or at the battery pack. PID (Photo Ionization detector) sensors detected most hydro carbons at a reasonable concentration (ppm-range) while IR detectors require higher concentrations. The FLIR thermal camera allows to detect increased temperatures but needs to be positioned towards "the problem area" (Martinsen et al., 2020). For detecting temperature changes, that can develop into a dangerous situation, it might not be a solution to attach several temperature sensors on a mining vehicle positing towards each possible problem area. Especially considering installation costs and the need for maintenance they might require. Therefore, drone inspections, with FLIR camera as a payload, has been tested at the open pit, in this research project to support early indication of increased temperature. In an underground mine the FLIR cameras could for instance be attached at the entrance and study the mining vehicles when they are going in and out of the mine. Combining machine data with signals from the sensors forms the basis for RCA and the development of decision trees. Covid-19 has forced us to postpone a number of test occasions at the Aitik site meaning some test data and results are missing. So, this paper is focusing on the methodology of decision trees.

3 Results with discussions

3.1 Tests occasions

A few drone test series have been carried out, initially to create an understanding of the drone's capacity with different payloads. The first test series concerned measuring of gases after blasting in normal weather conditions. The gas sensor attached to the drone had the functionality to measure multiple gases CO, NO and NO_2 and was of the same brand as the handheld sensor. This is a mandatory activity and according to the Swedish authorities (AFS, 2018) which has recently been updated, the gas levels it allows employees to be exposed for are listed in table 2. KGV is recommended maximum exposure value calculated as a time-weighted average over a reference period of 15 minutes. While NGV is the occupational exposure limits for one working day, normally 8 hours.

Table 2. Gas level according to the Swedish regulation.

Gases	Level (NGV) ppm	Level (KGV) ppm
CO	20	100
NO	2	0
NO_2	0,5	1

This activity is performed manually with a handheld sensor by the person in a vehicle who makes the first trip to the bottom of the pit, which takes about approximately thirty minutes to reach. Which is more time than the drone needed to measure the entire blasting area. If the level is not acceptable, they return to the top and wait a while before transporting themselves down to the pit to repeat the measurement. Not until acceptable levels are measured, the work can start to load the trucks with ore. The test observed that the drone was able to stay in the blasting area until acceptable gas level was measured. In Table 3, a fraction of all tests data is shown. The drone was situated at the same coordinates as the vehicle was driving to ensure we measured at the same place and at the same time.

Table 3. Some sample of gas measurement, manually vs drone mounted sensor in the north blasted area

Sensor	Handheld, ppm	Drone mounted, ppm
CO	6	6
NO	0	2,2
NO_2	0,04	0,14

The second test series was conducted during winter with a temperature of minus 35 °C. While third series of tests was conducted in heavy winds and rain. The drones used in the test showed almost problem-free regardless of weather conditions. Except in the very cold conditions when the wings froze, and we were forced to defrost. In Table 4, fraction of test data flying over the blasted area are shown. The indication is that measuring with a handheld sensor from the "road" will not capture the entire gas level situation.

Table 4. Some sample gas measurement, drone flight over blasted area, north.

Sensor	Drone mounted, ppm
CO	84
NO	8,5
NO₂	0,18

All three test sessions demonstrated the benefits of performing this type of work with a drone such as:
- The drone could fly over the whole blasting area compared to the vehicle that only could travel on the road, if it was navigable.
- Received more accurate data
- Replacing the vehicle with a drone will result in less CO_2 and is thus more environmentally friendly.
- Inhalation of these gases can be dangerous for humans. It can block the transport of oxygen to the brain and you develop an irreversible medical condition. Some cases have gone to court that ended in conciliation. The tests show that it will be safer to carry out the measurement with a drone equipped with gas sensors.
- If the blasting map is fed to the drone, it can fly in an autonomous (Chen, 2018) manner over the exact area.
- Regardless of the gas levels, the drone can remain in the area and measure and record continuously.
- Replacing this manual task with a drone allows savings of up to one-man week per year.

Further, the FLIR camera mounted on the drone was tested. The FLIR camera measures "hot spots", the highest temperature in the image, and what this temperature is. It indicates where the risk of, for example, fire is greatest. With this it is possible to scan and find risk temperatures in a sophisticated way like in Figure 10, where leakage was detected during the flights. The temperature itself may not say much more than possibly elevated "hot spot" values, but if these data are correlated with machine data and the values from the

gas data from the sensors mounted on the mining vehicle, it increases the chance of finding the fault that causes the situation. RCA with the support of decision trees (like e.g. in Figure 13-15) provide the opportunity for corrective measures to be proposed.

Figure 10. Leaking mining vehicle.

With a FLIR camera mounted on a drone, it supports inspections by scanning all mining equipment in the open pit area and allows it to be carried out on a more regular basis than today´s. On a drilling rig, for example, it is interesting to follow temperatures on engines, hydraulics and oil leaks, etc. On a conveyor belt the test indicated possibilities for detection on e.g. worn bearings. For an underground mining the same camera could be mounted at the entrance and exit to the mine and measure the temperature of the brakes as for tires. With image and pattern recognition comparing with previous pictures increased temperature values could be detected and warn at an early stage. In addition, tests with drones were conducted in the flotation area, in the crusher and finally inspection of the electrified Trolley Line shown in Figure 11 – 13.

Figure 11. Flotation area, rusty pipe.

Figure 12. Trolley Line objects.

Figure 13. Crusher.

From the described tests the following benefits could be observed:

- Possible to inspect areas where safety of the employee cannot be guaranteed
- Reach areas that had never been inspected before
- Time saving
- The images that the camera took from the drone flights was possible to be viewed by several employees in a superior surveillance system room simultaneous, and thus support and streamlining the troubleshooting process when more than two eyes look at the problem.

3.2 Decision Trees

The above conducted tests show possibilities with both sensors and drones in the mining business. Increasing production will require several solutions that allow equal productivity to be maintained or even further improved. Using only drones and sensors is not the whole solution. It will be necessary to ensure that all collected data and images are processed in an efficient manner and communicated to an overall production and safety system so that warnings and alarms can be acted upon. A piece of the puzzle for transforming into autonomous mining operation decision trees for each aspect in the mining process is proposed to support monitoring and perceive changes in production more efficiently and for minimizing productions stops for maintenance. Some examples of decision trees can be seen in Figures 14-16.
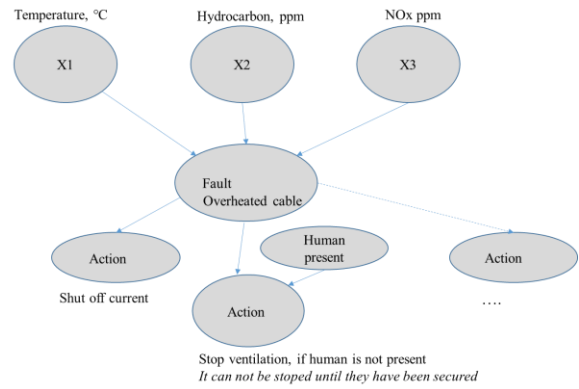
Figure 14. Decision tree for overheated cable.
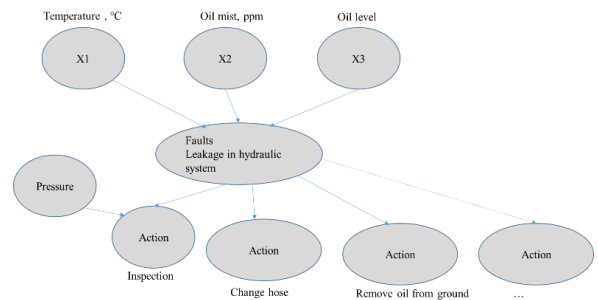
Figure 15. Decision tree for leakage in hydraulic system.
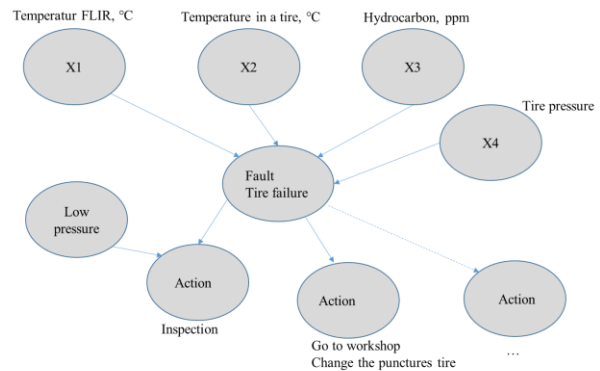
Figure 16. Decision tree for tire failure.

4 Conclusions

Future mining production is predicted to develop into fully autonomous mining processes. Safety and security will be important aspects in such operation. To be able and catch early changes in production predictive and proactive maintenance activities is preferred. Detecting changes in production that may cause a stop needs to be quicker than today. This requires sustainable solutions supported with smart inspections and early discovery of changes before the situation develops into a dangerous state. Today's maintenance inspections consist been verified that some of them could be supported with

drones equipped with appropriate payloads. A compilation of the results from the drone tests have been presented including possible influence of the investigated parameters a decision trees methodology was introduced.

The results show that drones equipped with different payloads can support the conversion to succeed with autonomous mining concept and be the enabler for safer and efficient work methods, inspect parts in hard-to-reach areas or not reachable for a human at all. But not least the possibility of reducing carbon dioxide emissions from transport by mining vehicles, powered by diesel, for inspections when they can be advantageously replaced by drones, possibly powered by batteries. Which in turn contributes to the goal that Sweden succeeds with the goal of becoming a fossil-free country by year 2045.

A few decision trees were suggested in order to support the future development of an AI application that automatically suggests maintenance actions. The decision tree contributes to increased safety by minimizing time for conducting maintenance work in hard to reach areas and prevent dangerous situations from occurring. If the drone inspections can be initiated and monitored automatically via a central operating system and using autonomous routes, the conditions for a self-running inspection process can be created.

Future research activities will focus on diagnostic supported by deep learning, pattern recognition as for artificial intelligence applications. In addition, autonomous inspection routes will be evaluated. The authors believe that the use of machine learning and artificial intelligent application creates the conditions for reducing acute production stoppages and facilitates maintenance inspections and thus creates the conditions for productivity to be maintained or even more likely to improve as the production increases.

Acknowledgements

References

A. Abraham et al. (Eds.), Soft Computing Systems - Design, Management and Applications, pp. 221-230. IOS Press, Amsterdam, the Netherlands, 2002.

Khajanchi Amit. *Artificial Neural Networks. The next intelligence*, 2003.

AFS. Arbetsmiljöverket. Hygienska gränsvärden; 2018:1

Michael Berthold and David J. Hand. *Intelligent Data Analysis an introduction.* Spinger-Verlag Berlin Heidelberg New York: Publishing, 1999.

Shihui Chen, Umut Durak, and Sven Hartmann. Modeling and Simulation-based Development of Autonomy Features for Drones, *SNE* 28(2):55-60, 2018. doi: 10.11128/sne.28.tn.10413.

M. W. Craven and J. W. Shavlik. Using neural networks for data mining. *Future Generation Computer Systems,* 13:211–229, 1997.

W. S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:115-133, 1943.

The Drone Market Report 2019-2024. Drone Insight Industry

Rickard Hansen. *Design fires in underground hard rock mines*, School of Sustainable Development of Society and Technology, Mälardalens University, Licentiate Thesis 127, Västerås, Sweden, 2011.

Rickard Hansen. *Study of heat release rates of mining vehicles in underground hard rock mines*, Mälardalens University Press PhD Dissertation No. 178, Västerås, Sweden, 2015.

Rickard Hansen. *Fire statistic from the mining industry in New South Wales, Queensland and Western Australia*, The University of Queensland. doi:0.131340/RG.2.2.22666.777 67, Australia, 2018.

S. Haykin. *Neural Networks*, Prentice Hall, 1999.

Dietmar Kasper, Galia Weidl, Thao Dang, Gabi Breuel, Andreas Tanmke, and Andreas Wedel. Object-Oriented Bayesian Network for Detection of Lane Change Maneuvers. *IEEE Intelligent Transportation System Magazine*, 2012.

Madeleine Martinsen, Erik Dahlquist, Anders Lönnermark and Örjan Säker. Gas sensors for early detection of fire hazards caused by vehicles in underground mines. *International Conference on Scandinavia Society, SIMS 2019*. Aug 13-16, 2019, Västerås, Sweden. Paper ID: 28.

Madeleine Martinsen, Erik Dahlquist, Anders Lönnermark, and Örjan Säker. Energy and Safety Diagnostic in Underground Facilities. *Conference paper ISTSS Munich, March 2020*.

Yichang (James) Tsai, Using Image Pattern Recognition Algorithms for Pressing Video Log Images to Enhance Roadway Infrastructure Data Collection. Final Report for Highway IDEA Project 121, 2009.

G. Weidl, A.L. Madsen, and S. Israelson, Applications of object-oriented Bayesian networks for condition monitoring, root cause analysis and decision support on operation of complex continuous processes, *Computers and Chemical Engineering*, 29: 1996-2009, 2005.

G. Weidl, A. L. Madsen, and E. Dahlquist. *Condition monitoring, Root Cause Analysis and Decision Support on Urgency of Actions.* In Book Series FAIA (Frontiers in Artificial Intelligence and Applications), vol. 87.

G. Weidl, A. Madsen, and E. Dahlquist. Decision support on complex industrial process operations. pages 313-328, In O. Pourret, P. Naim, and B. Marcot, editors, *Bayesian Networks, a practical guide for applications.* John Wiley. ISBN 978-0-470-06030-8, 2008.

J. M. Zurada. *An introduction to artificial neural networks systems*. St. Paul: West Publishing, 1992.

Front Tracking of Shock and Combustion Waves by an Optimal Transport Framework

Sabin Bhattarai   Ola Marius Lysaker   Dag Bjerketvedt

Department of Process, Energy and Environment Technology, University of South-Eastern Norway, Norway,
`{sabinbhattarai1994@gmail.no, marius.lysaker@usn.no, dagbj@usn.no}`

Abstract

This paper presents a framework for front tracking of shock waves from high-speed video recordings. The video was corrupted with noise and artifact that made front tracking a challenging task. To our knowledge, the implementation of Unnormalized Optimal Transport for determination of velocity is novel.

An Unnormalized Optimal Transport framework was implemented in Python to determine the route of propagation from the recorded high-speed video. The obtained route was further computed for front tracking by two methods; Divergence and Transport method. These methods were investigated with both synthetic images and the recorded high-speed video frames. For both, the Transport method provided better results than the Divergence method. Therefore, the shock wave velocity and the shock angles were calculated from Unnormalized Optimal Transport combined with the Transport method. Preliminary results indicate that our findings are in good agreement with sensor-based measurements.

This framework verified that the triple point height is increasing, and the oblique shockwave moves faster than the normal shock wave.

Keywords: unnormalized optimal transport, front tracking, image processing

1 Introduction

Combustion and shock waves propagate at high speed which makes challenging to find the velocity and shape of the flame and shock wave. In lab experiments, state-of-the-art has been to use of sensors like pressure transducers to estimate velocities. One drawback of that technique is the limited number of sensors placed inside a test rig. An alternative to these sensor-based measurements is to use a high-speed camera whereon the position, shape, and velocity can be found from the recorded high-speed images. Video recording can be analyzed frame by frame to track wave. Even though humans can comfortably identify and track such waves in a video, it is challenging to design robust algorithms for automatic tracking. Also, the presence of noise and artifact made the task more challenging. Although various algorithms, such as Level set (Osher and Sethian, 1988), Splines (Maharjan et al., 2019), etc. have been proposed, we are evaluating unnormalized optimal transport for tracking waves.

The French mathematician Gaspard Monge pioneered on optimal transport in 1781. Later, the Russian Mathematician Leonid Vitaliyevich Kantorovich stated and proved the problem through the additional functional tools known as duality theorem. Monge explained optimal transport as distance covered to transfer a pile of sand from one location to another with minimum transport cost. If this cost is a function of distance, the distance from optimal transport is called the Kantorovich-Rubinstein distance or Wasserstein distance (Villani, 2008). Optimal transport gained popularity in inverse problems and machine learning.

The above classical approach from Monge and Kantrovich is in the normalized density space, i.e., the masses of two density functions or histogram is equal. Gangbo et al. (2019) have described a formulation in which the masses of two density functions are unequal, refer to it as unnormalized or unbalanced optimal transport. With this approach, it is possible to examine two images with different intensities. Therefore, the high-speed video was analyzed with unnormalized optimal transport.
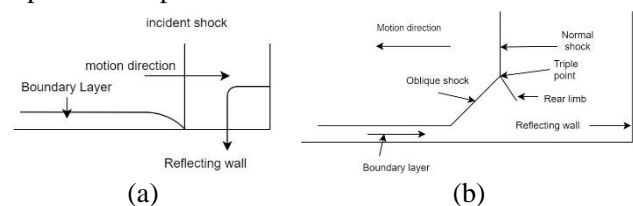
Figure 1. Schematic diagram of bifurcated shock wave a) Incident shock wave b) Reflected shock wave

The high-speed video images provided the shock wave position. A shock wave is a thin compression wave across which the gas changes its thermodynamic properties such as density and pressure (Bjerketvedt et al., 1997). When a normal incident shock wave hits the end wall, it is reflected. The reflected shock wave is bifurcated due to the boundary layer. Figure 1 shows the bifurcation of the reflected shock wave. This occurs if stagnation pressure in the boundary layer is less than

the pressure behind the normal reflected shock wave (Damazo et al., 2012).

2 Methods

The Test rig of the experiment is described in Siljan et al. (2017). The high-speed video from this experiment was analyzed with methods below:

2.1 Image conversion

Image conversion is pre-processing for optimal transport. The input for optimal transport is in densities space. The pixel values were converted into densities and must be non-negative values. The obtained images were resizing into the same scale.

2.2 Unnormalized Optimal Transport

The unnormalized optimal transport was implemented to determine the route of the traveling wave. L^1-norm was used as a distance function in the unnormalized optimal transport as it requires less computation time and is simple for implementation compared to the L^2-norm (Li et al., 2018). So, implemented optimal transport is L^1- unnormalized optimal transport. The algorithm for L^1- unnormalized optimal transport is mention below (Osher, 2019):

$$m^{k+1} = \frac{m^k + \tau_1 \cdot \nabla \Phi}{|m^k + \tau_1 \cdot \nabla \Phi|}(|m^k + \tau_1 \cdot \nabla \Phi| - \tau_1) \quad (1)$$

$$\Phi^{k+1} = \Phi^k + \tau_2(\nabla \cdot m^{k+1} - \nabla \cdot m^k + \mu^1 - \mu^0 + c). \quad (2)$$

Where,
$c = \frac{1}{row \times col}(sum\ of\ \mu^1 - sum\ of\ \mu^0)$,
$row = number\ of\ rows\ of\ \mu^1 or\ \mu^0$,
$col = number\ of\ columns\ of\ \mu^1 or\ \mu^0$.
Here, μ^1 and μ^0 are input images that display the movement of an object. τ_1 and τ_2 are step sizes. c is constant for compensation of change of intensities between images.

The solution obtained is m that is a vector, and Φ is the Lagrange multiplier. The m consists of two components represented as m_x and m_y. It was found that m_x and m_y are responsible for the horizontal movement and vertical movement, respectively.

Initially, the value of m and Φ is zero. Iteration continued until the solution converged. The iteration time depends upon grid size or pixel size of images, i.e., smaller grid sizes mean more data that increases the execution time. So, multilevel unnormalized transport is implemented for faster calculation (Liu et al., 2019).

A detailed study of this method can be found in (Gangbo et al., 2019; Chambolle and Pock, 2011; Li et al., 2018).

2.3 Post-processing

The outcomes from unnormalized optimal transport are used as input to the front tracking. This process is accomplished by using the Transport and Divergence method.

2.3.1 Transport method

The resultant from these two components are calculated using the formula below:

$$resultm = \sqrt[2]{m_x^2 + m_y^2}. \quad (3)$$

The obtained result has proceeded for edge detection after Total Variation (TV) denoising. TV denoising is used to filter $resultm$ (from above equation) as it preserves the edge better (Chambolle and Pock, 2011). Edge detection is accomplished by the Sobel Operator (Bovik, 2009). The maximum of each row from the initial column to the middle of the unnormalized optimal transport route is used as a tracked front from edge detected images.
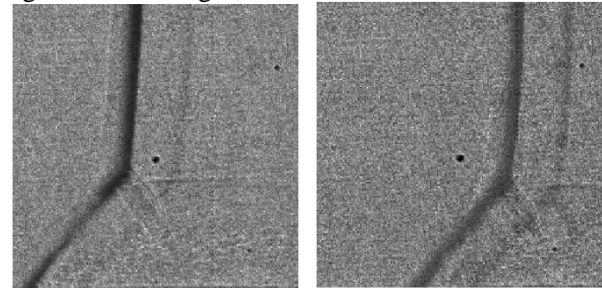
Figure 2. Images presenting the propagation of shock wave.

(a)   (b)
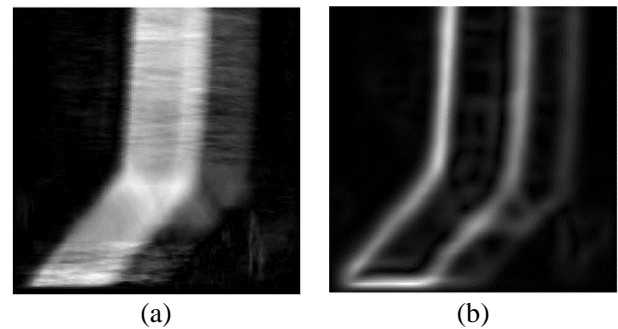
Figure 3. a) Unnormalized optimal transport outcome b) Edge detection of a)

Consider Figure 2 as inputs for unnormalized optimal transport to demonstrate this method. The obtained solution m is formulated with (3) is in Figure 3(a). This figure undergoes TV denoising due to noise caused by turbulence from the shock wave movement. The denoised result is used for edge detection with the Sobel operator in Figure 3(b). The data extracted as maximum from the left edge from Figure 3(b) as the right edge is disturbed by the rear limb in Figure 1(b). The tracked front is in Figure 4.
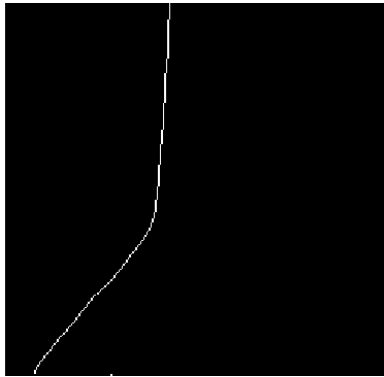
Figure 4. Tracked front from Figure 2

2.3.2 Divergence method

The dynamic constraint for unnormalized optimal transport is integrated with the time variable (Gangbo et al., 2019). We obtain the following equation:

$$\nabla \cdot m = \mu_0 - \mu_1 + c. \tag{4}$$

Here, the initial image ($\boldsymbol{\mu_0}$) refers to an object's initial position in an image, and the final image ($\boldsymbol{\mu_1}$) refers to the moved position object in the next one during movement. The divergence of m incorporates initial image and final images, and $\boldsymbol{c}$ is constant to deal with intensities change. If the final images ($\boldsymbol{\mu_1}$) is zero, then the remaining will only be initial images and difference between initial and final images intensities, i.e., $\boldsymbol{\nabla \cdot m = \mu_0 + c}$. So, image from $\boldsymbol{\nabla \cdot m}$ is used for front tracking. This method is illustrated using the same example as above.

(a)   (b)
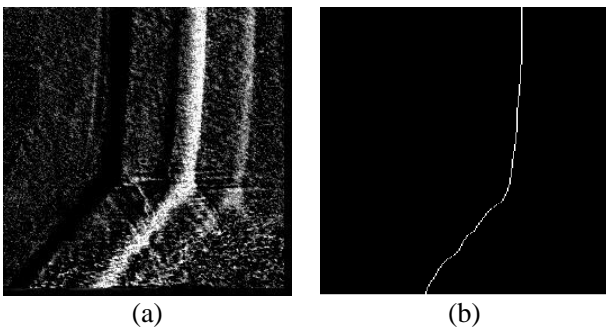
Figure 5. a) Divergence of unnormalized optimal transport result(m) b) Tracked front from Figure 5(a)

Figure 5(a) is the divergence of m with $\mu_1 = 0$. The obtained image is similar to the second image in Figure 2, but the shock wave is more distinct. The maximum of each row is extracted and presented in Figure 5(b).

3 Result and Discussion

3.1 Comparing methods

The unnormalized optimal transport outcome was further analyzed to calculate velocity by using the two methods mentioned above. These two methods are compared with noisy and noise-free synthetic images and experimental video images below.

3.1.1 Synthetic images

The developed series of images displays a slender's horizontal movement in Figure 6 as the shock wave movement in the video.
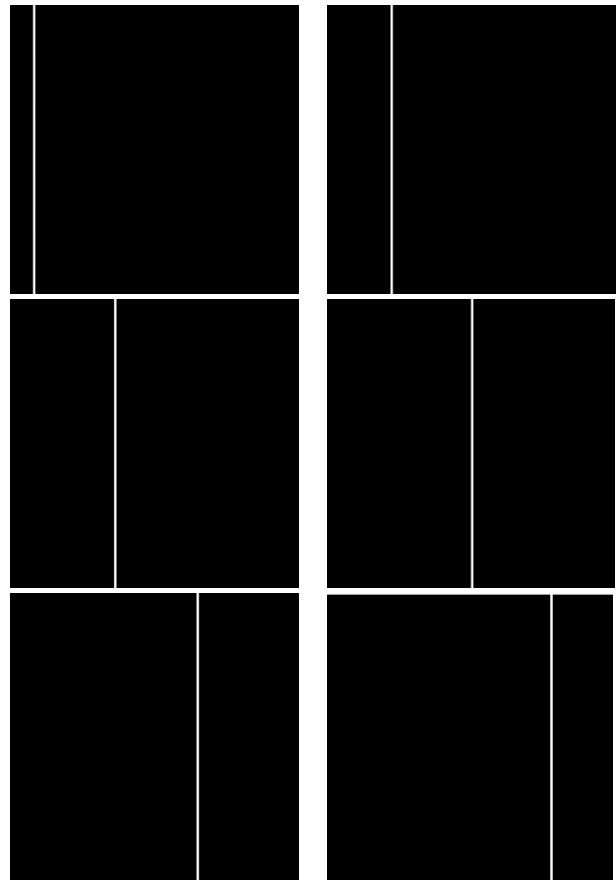
Figure 6. Series of the synthetic image showing slight movement.

The transport and divergence method applied to the solution of unnormalized optimal transport shown in Figure 6. The results from these methods are in Figure 7(a) and Figure 7(b), respectively. The transport method results in Figure 7(a) tracked slender more accurately than the divergence method results in Figure 7(b). Even though the transport method shows a better result, it cannot find the first row of each slender's position in images.

(a)   (b)
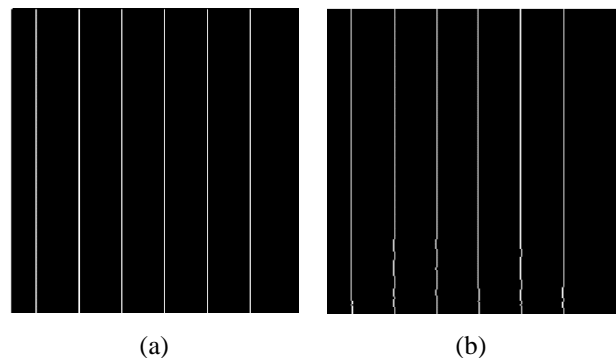
Figure 7. Tracked slender from noise-free images a) Transport method b) Divergence method

The images from Figure 6 are added with noise. Noise is a normal distribution whose mean is zero and the standard deviation is 10. The remaining assumption is of these images are equivalent to noise-free synthetic images. The slender is tracked based on the above mention methods, illustrated in Figure 8. The Transport method cannot determine the front of the first and last row, but the remaining portion of the tracked slender shown in Figure 8(a) is exact as input. The determined slender by divergence method is zigzag shown in Figure 8(b). The transport method is better than the divergence method displayed in Figure 8.

(a)   (b)
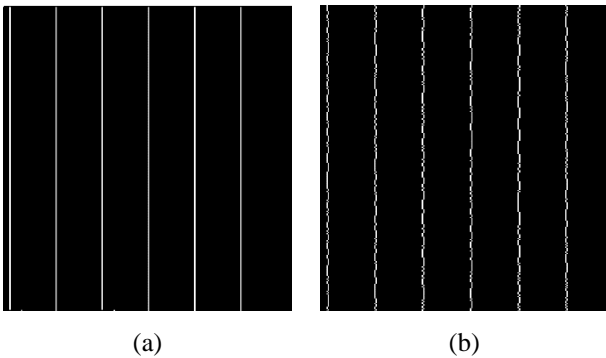
Figure 8. Tracked slender from noisy images a) Transport method b) Divergence method.

3.1.2 High-speed video

The high-speed video from the experiment includes a series of images displaying the shock wave movement, i.e., incident shock and reflected shock wave. The video was broken into images, and a single image refers to the frame. Images from frame number 80 to frame number 175 of video are input for analysis. The unnormalized optimal transport was implemented at an interval of 5 frames. The outcome of unnormalized optimal transport is the data for tracking front using the divergence and transport method illustrated in Figure 9.

Visually, the front tracking with the transport method is better than the divergence method from Figure 9. Although the tracked fronts are bend at the bottom with the transport method, it has a smoother transition between tracked points for wave than the divergence method. The problem might be the data extraction procedure in the divergence method. So, the results are favoring the transport method. Hence, we are proceeding with the transport method.

(a)

(b)
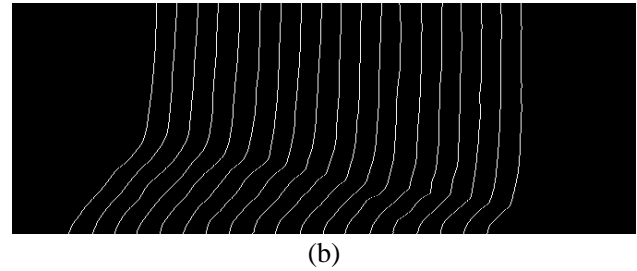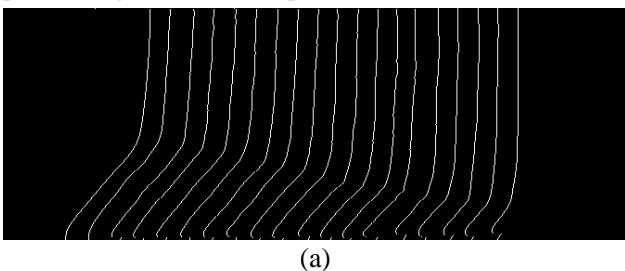
Figure 9. Tracked front from reflected shock wave with unnormalized optimal transport at an interval of 5 frames a) Transport method b) Divergence method.

The unnormalized optimal transport was implemented at an interval of 3 frames, and shock wave tracking accomplished using the transport method as illustrated in Figure 10. In comparison between Figure 9(a) and Figure 10, it was found that unnormalized optimal transport at the interval of 5 frames is providing better results than at the intervals of 3 frames. The further analysis uses the unnormalized optimal transport at the intervals of 5 frames and the transport method.
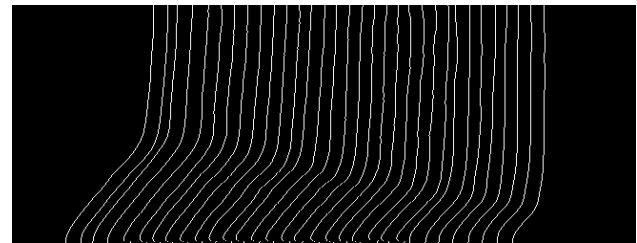
Figure 10. Tracked front from reflected shock wave with unnormalized optimal transport at an interval of 3 frames and Transport method

3.2 Velocity Determination

Assume two fronts shown in Figure 11 for the calculation of the velocity of the shockwave. Let us denote $d_1(x, y)$ as a pixel position on the front (f) and $d_2(x, y)$ as a pixel position on the consecutive front ($f + 1$) shown in Figure 9. Here, x and y represent row and column in images as per Python.
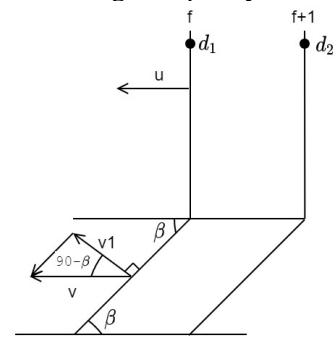
Figure 11. Schematic sketch of two consecutive shock waves.

The distance between the fronts is calculated as $d = |d_1 - d_2|$. The obtained difference undergoes mathematical operation of the scaling factor and the high-speed camera frame rate for computation of actual

velocity. The scaling factor for an image from the high-speed video is $1\ mm = 9.528 \pm 0.079\ pixel$. The frame rate of the high-speed camera is 500000 frames per second (FPS). The velocity is calculated with (5) for images at the interval of 5 frames.

$$v = \frac{d}{5} \times \frac{500000}{9528}\ [\frac{m}{s}] \qquad (5)$$

The velocity (u) represented the normal shock wave, and (v) represented the oblique shock wave. The oblique shock wave is tilted with angle (β), which is known as the shock angle. The oblique shock wave velocity is divided into normal and parallel components. The normal component of velocity (v) is represented as $v1$.

$$v1 = v \cdot \cos(90 - \beta) = v \cdot \sin\beta \qquad (6)$$

The 20 points from the upper portion of each tracked fronts from row number 1 to row number 21 were used to calculate the normal shock wave's average velocity. Similarly, 20 points starting from row number 195 to 215 are extracted to calculate the oblique shock wave's average velocity. The normal velocity for the oblique shock wave can only be calculated after the calculation of shock angles.

Figure 12, oblique shock wave crossed by the upper and lower boundary, is data for estimating shock angles. This figure illustrated that the upper and lower boundary crosses tracked oblique shock. The lower boundary is straight, but the upper boundary is not because the oblique shock length increases (the elevation of the triple point is increasing) as shock wave propagates. The data is extracted from the upper to lower boundary for each front separately. The extracted data trend is a straight line, and the straight line equation is fitted with extracted data by the least-squares method. The inclination of the obtained straight-line reference to the bottom side of images is shock angles. These angles and oblique shock wave velocities are required to calculate the oblique shock wave's normal velocity with (6).
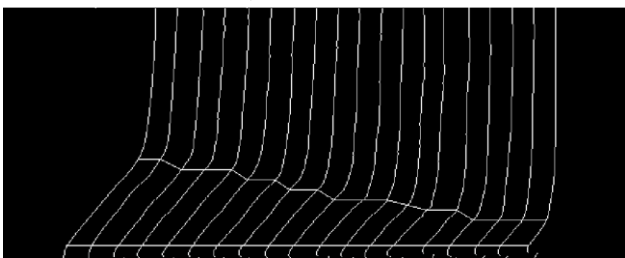
Figure 12. The region selected for calculation of shock angles from tracked fronts with transport method

The velocities calculated above are presented in Figure 13. The x-axis in this figure is the distance from the shock reflected wall. It is estimated from 10 points from upper tracked fronts.
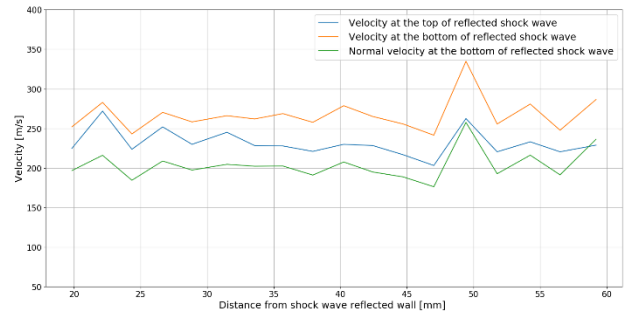
Figure 13. Average velocities of normal and oblique shockwave, and normal component of an oblique shock wave with unnormalized optimal transport with 5 frames and transport method.

Figure 13 displays that the oblique shock wave moves faster than a normal shock wave as in the high-speed video. This result is obtained without filtering inputs for unnormalized optimal transport.
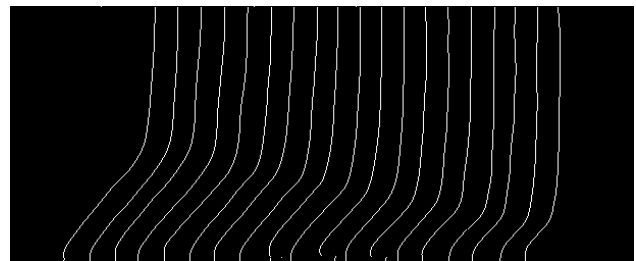
**Figure 14.** Tracked fronts from reflected shock wave with unnormalized optimal transport at an interval of 5 frames in which series of operations performed in input images for denoising and transport method.

The images from the given high-speed video are noisy. Hence, images are denoised before the implementation of unnormalized optimal transport. In this process, background subtraction, morphological operation (closing), and TV denoising proceed in a sequence. Firstly, the first image from the given video subtracted from noisy images displaying shock propagation. Successively, the morphological operation (closing) is performed in subtracted images as it fills a tiny gap near the edge (Maharjan et al., 2019) followed by TV denoising. The obtained images are input for unnormalized optimal transport followed by the transport method, and the result in Figure 14.

Figure 14 exhibits a smoother curve of shock wave than in Figure 9(a) but there is not much change in tracked front with filtered and unfiltered inputs for unnormalized optimal transport. The velocities calculated for filter images from (5) and (6) are in Figure 15. Figure 13 shows higher oscillation in the velocities curve than in Figure 15 as denoising operations in input images. Although result from unfiltered images produces high oscillation velocities curves, the trend for normal shock wave velocity, oblique shock wave velocity, and normal velocity for filter and unfiltered input of unnormalized optimal transport similar.
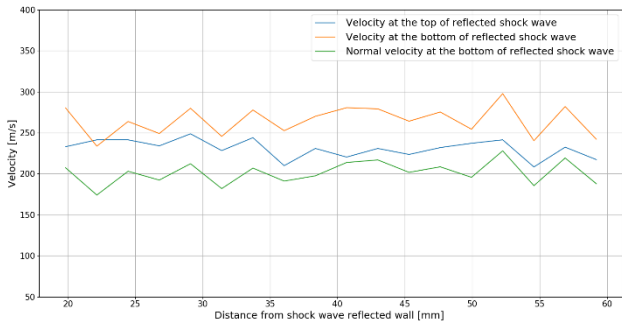
Figure 15. Average velocities of the normal and oblique shock wave, and normal component of the oblique shock wave with unnormalized optimal transport at an interval of 5 frames in which series of operation performs in input images denoising and transport method.

Figure 16 demonstrates that trends of shock angles of unfiltered and filtered images are slightly different but fluctuate around 50°. The average shock angles of unfiltered and filtered images are 49.7° and 49.4°, respectively.
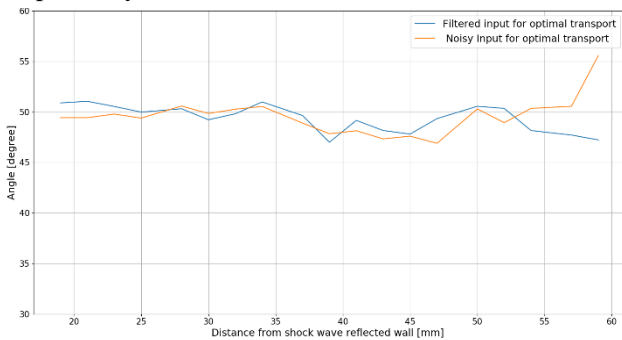
Figure 16. Oblique shock angle from unnormalized optimal transport at an interval of 5 frames and transport method.

If a tracked front deviate from one pixel, then it results in an error of $11\ m/s$ considering images as an input interval of 5 frames.

3.3 Comparison with combustion, process safety, and explosion research group

Researches had conducted simulation for computation of the velocity with the same high-speed video. The solution from pattern matching and segmented regression (Siljan et al., 2017), dynamic template matching (Maharjan, 2019), and snake model with gradient vector fields and watershed segmentation (Maharjan et al., 2019) and sensor-based calculations are in Table 1. The mentioned frameworks and experimental results for oblique shock wave velocities are within 5 m/s deviation but above 10m/s deviation for normal shock wave velocities. The pressure sensors are presented at the bottom, so the normal shock wave velocities are calculated from the oblique shock wave pressure.

Similarly, the difference in shock angle between simulated results and experimental value is less than

3º. The results of unnormalized optimal transport and other frameworks are within 5 % fluctuation.

Table 1. Comparison between results from different methods for reflected shock wave

S.N.	Description	Average velocity of normal shock wave[m/s]	Average velocity of oblique shock wave[m/s]	Average shock angle
1.	Result from the pressure transducer (Maharjan et al., 2019)	216	266	48º
2.	Result from template matching (Siljan et al., 2017)	230	262	50º
3.	Result from snakes model (Maharjan et al., 2019)	227	267	48.9º
4.	Result of optimal transport with filter images (here)	230.8	265	49.4º
5.	Result of optimal transport with unfiltered images (here)	231.5	267	49.7º

4 Conclusion

The above section demonstrated that the present framework can estimate velocities and shock angles from experimental results and synthetic images from other algorithms. The velocity is obtained from a series

of images representing an object's motion with unnormalized optimal transport.

The transport method provided better front tracking of unnormalized optimal transport results than the divergence method. Moreover, filtering the input to the unnormalized optimal transport had only minor impact on the results.

5 References

A. Bovik. *The Essential Guide to Image Processing*. Elsevier Science & Technology, 2009. Url:http://ebookcentral.proquest.com/lib/ucsn-ebooks/detail.action?docID=452947.

A. Chambolle, and T. Pock. A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging. *Journal of Mathematical Imaging and Vision*, *40*(1):120–145, 2011. doi: 10.1007/s10851-010-0251-1.

W. Gangbo, W. Li, S. Osher, and M. Puthawala. Unnormalized Optimal Transport. *Journal of Computational Physics*, *399*, 108940, 2019. doi: 10.1016/j.jcp.2019.108940.

W. Li, E. K. Ryu, S. Osher, W. Yin, and W. Gangbo. Parallel Method for Earth Mover's Distance. *Journal of Scientific Computing*, *75*(1): 182–197, 2018. doi: 10.1007/s10915-017-0529-1.

J. Liu, W. Yin, W. Li, and Y. T. Chow. Multilevel Optimal Transport: A Fast Approximation of Wasserstein-1 distances, 2019. doi: http://arxiv.org/abs/1810.00118.

S. Maharjan, D. Bjerketvedt, and O. M. Lysaker. An Image Processing Framework for Automatic Tracking of Wave Fronts and Estimation of Wave Front Velocity for a Gas Experiment. In L. Chen, B. Ben Amor, & F. Ghorbel (Eds.), *Representations, Analysis and Recognition of Shape and Motion from Imaging Data*: 45–55, 2019. doi: 10.1007/978-3-030-19816-9_4.

S. Osher. *Unnormalized Optimal Transport*.,2019. doi: https://ipam.wistia.com/medias/2015gis083?wtime=0.

S. Osher and J.A. Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations. *Journal of Computational Physics*, *79*(1): 12–49, 1988. doi: 10.1016/0021-9991(88)90002-2.

E. Siljan, M. Lysaker and S. Maharjan. *Wave Front Tracking using Template Matching and Segmented Regression*: 326–331, 2017. doi: 10.3384/ecp17138326.

Anaerobic Digestion of Hemicellulosic Sugars Implemented in ADM1

Vibeke Bredvold Karlsen   Nirmal Ghimire   Rune Bakke   Wenche Hennie Bergland

Department of Process, Energy and Environmental Technology, University of South-Eastern Norway

Vibeke.B.Karlsen@usn.no, Nirmal.Ghimire@usn.no, Rune.Bakke@usn.no, Wenche.Bergland@usn.no

Abstract

Lignocellulosic biomass is a sustainable and renewable source for both solids like biochar and biomethane by anaerobic digestion (AD). Hot water extraction (HWE) improves the total utilization of the biomass and produces a hydrolysate rich in hemicellulosic sugars with characteristics that needs to be understood. A study of the AD process in batch reactors with synthetic substrates composed of hemicellulosic sugars was performed and modelled using the standard IWA Anaerobic Digestion Model No. 1 (ADM1). Simulations were also performed using three strategies: 1) varying the stoichiometry for monosaccharide degradation, 2) varying the maximum uptake rate of the monosaccharide degrading organisms and 3) including a first order rate limiting step. The ADM1 model is a good tool but gave initially moderate agreement with the experimental results. The first two strategies did not improve the simulations but it improved significantly upon incorporating a rate limiting step, thereby simulating possible effects based on types of microorganisms present and diffusion limitations.

Keywords: anaerobic digestion, lignocellulosic hydrolysate, hemicellulosic sugars, ADM1, first order step

1 Introduction

Renewable materials can replace fossil materials as a source for energy and to produce new and existing products. This has shifted the focus to renewable material sources. Lignocellulosic biomass, such as woody biomass, agricultural and forest residues, energy crops and grasses, is a large potential source of new materials and renewable energy. Treatments such as hot water extraction and steam explosion, commonly used on lignocellulosic biomass to remove volatiles before renewable solid material production, extract hemicellulose from the lignocellulosic biomass. The hemicellulose shows promising results as a degradable material in anaerobic digestion for biogas production, whereas untreated lignocellulosic biomass has low biodegradability (Hu and Ragauskas, 2012) due to the main components cellulose (38-50 %), hemicellulose (23-32 %) and lignin (10-25 %) being linked together

and packed closely. Through hydrolysis of hemicellulose during hot water treatment, easily degradable pentose (xylose and arabinose) and hexose sugars (glucose, mannose and galactose) are produced along with acetic acid (Amidon and Liu, 2009).

The concentration of sugars varies in lignocellulosic biomass. Softwoods, like spruce and pine, contain more mannose and glucose in comparison with the other sugars while hardwoods, like aspen, oak and willow, contain a higher amount of xylose (Palmqvist and Hahn-Haagerdal, 2000).

The Anaerobic Digestion Model No.1 (ADM1) (Batstone et al., 2002a) has been applied for different AD systems and its performances studied for various substrates and reactor configurations (Gehring et al., 2013). The ADM1 model is developed to handle sludge with complex materials consisting of fat, protein and carbohydrates. However, simulation of AD of individual hemicellulosic sugars found in lignocellulosic biomass in ADM1 is still rare and needs to be studied carefully to understand underlying mechanisms to increase the biodegradability. Processes related to carbohydrates are important for lignocellulosic biomass.

The aim of this paper is to evaluate ADM1 as a tool for simulating AD of sugars typically found in hydrolysates prepared by hot water extraction of lignocellulosic biomass. Kinetic parameters were estimated based on batch experiments for each of the sugars which were made synthetically. Moreover, simulations were performed with variations in the stoichiometry during VFA production from monosaccharides and the soluble sugar degrading kinetic values and compared with the experimental results.

2 Materials and Methods

Major hemicellulosic sugars present in hydrolysate of lignocellulosic substrates, but made synthetically here, are tested in anaerobic batch reactors and modelled in ADM1 individually.

2.1 Material Characterization

2.1.1 Substrate

Five different synthetic substrates were prepared from powder of one of the hemicellulosic sugars glucose, mannose, arabinose, galactose and xylose (Table 1). The D-form of all the sugars were used in the experiments. The D-form of the sugars are most common in nature, except for arabinose most commonly present in L-form (McMurry, 2014).

Table 1. Sugars used to produce the synthetic substrates

Sugar (Sigma-Aldrich)	Mass percentage of sugar
D-Glucose	≥ 99.5 %
D-Mannose	≥ 99 %
D-Xylose	≥ 99 %
D-Arabinose	≥ 98 %
D-Galactose	≥ 99 %

Each sample was prepared by diluting 1 gram of the respective sugar with 100 mL of distilled water to a sugar concentration of 10 g L^{-1}. Macronutrients (Table 2) and micronutrients (Table 3) were added to provide required nutrients with a COD: N: P ratio of 350:5:1 (Baeta et al., 2013). NaOH was added to increase the pH to a range of 7-7.5.

Table 2. Composition and concentration of macronutrients

Macronutrients	
Type of chemical	Concentration (g L^{-1})
NH$_4$Cl	44.48
(NH$_4$)H$_2$PO$_4$	5.3
(NH$_4$)$_2$HPO$_4$	1.78
MgCl$_2$·6H$_2$O	21.4
CaCl$_2$·2H$_2$O	7.56
NaHCO$_3$	100

Table 3. Composition and concentration of micronutrients

Micronutrients	
Type of chemical	Concentration (g L^{-1})
Yeast Extract	2.5
FeCl$_3$·6H$_2$O	0.2
ZnCl$_2$	5.2
MnCl$_2$·4H$_2$O	0.047
(NH$_4$)$_6$Mo$_7$O$_{24}$·4H$_2$O	0.064
AlKO$_8$S$_2$·12H$_2$O	0.01
CoCl$_2$·6H$_2$O	0.2
NiCl$_2$·6H$_2$O	0.52
H$_3$BO$_3$	0.12
CuCl$_2$·2H$_2$O	0.32
HCl	20

2.1.2 Inoculum

The inoculum used was granular sludge (Table 4) obtained from a mesophilic industrial internal recirculation reactor treating paper mill effluent.

Table 4. Properties of granular sludge

Parameters	Values
Density (kg m^{-3})	1.00 – 1.09
Diameter (mm)	0.6 – 2.7
Settling velocity (m h^{-1})	68 – 71
Total Solids (g L^{-1})	181
Volatile Solids (g L^{-1})	119
pH	7.46

2.2 Batch Reactors Set up

100 mL syringes were used as batch reactors in accordance with Ostgaard et al. (2017). Five parallel batch reactors were prepared for each sugar and fed with 15 mL inoculum and 14.1 mL substrate to have a uniform organic load of 10 gCOD L^{-1}. 3 blank parallel reactors containing only inoculum and distilled water were also prepared to correct for biogas generated from the inoculum. All the reactors were run at 35°C for 19 days.

2.3 Analytical Methods

Biogas production was measured manually in accordance with Ostgaard et al. (2017), and gas composition was measured when the volume was higher than 5 mL by gas chromatography (SRI 8610-C) as described by Bergland et al. (2015). Liquid phase COD (total and soluble), volatile fatty acids (VFAs), including acetate, propionate, butyrate, iso-butyrate, valerate, iso-valerate, iso-caprinate, caprinate and heptanate, pH and ammonium content (NH$_4^+$) were sampled and measured as described in Bergland et al. (2015).

2.4 Modelling and Simulation Methods

Simulations of batch tests were performed using the Anaerobic Digestion Model No. 1 (ADM1) in Aquasim. The ADM1 was applied to model the processes with stoichiometric coefficients, equilibrium coefficients, dynamic states and algebraic variables as proposed by Batstone et al. (2002b), for all biochemical and physio-chemical processes.

The following strategies were used to simulate batch reactors with hemicellulosic sugars in ADM1:

- Using kinetic and stoichiometric parameters as proposed by Batstone et al. (2002a).
- Varying the stoichiometry during VFA production from monosaccharides, affecting the acetate (fac,su), butyrate (fbu,su) and propionate (fpro,su) obtained through monosaccharide degradation.
- Varying the maximum uptake rate of the monosaccharide degrading organisms (km,su).

- Adding the sugars as carbohydrates, thereby including a first order step.

In all the simulations, the reactor volume was set to 29.1 mL and the temperature to 35°C in accordance with the batch experiments.

2.4.1 Substrate characteristics

Substrate concentrations implemented (Table 5) were calculated based on the substrate characteristics. Since the substrates only contained sugars (sCOD), the amount of sugar was added as S_{su}. In ADM1 the hexose glucose is used as a model monomer. Pentoses have similar yields as the hexoses, but obtain one less CO_2 or carboxylic acid (Batstone et al., 2002a). All the individual sugars were therefore included as S_{su} without modifying the stoichiometry of the sugars, and the batch reactors were added equal amounts of sCOD. The parameters related to particulates (X_c, X_{ch} and X_I) were set to zero in the simulations.

Table 5. Concentrations of hemicellulosic sugars in the substrates

Parameters	Description	Values used in the simulations
S_{su}	Monosaccharides	5.136 kgCOD m^{-3}
S_I	Soluble inerts	0.36 kgCOD m^{-3}
S_{IN}	Inorganic nitrogen	0.0092 M

The S_{cat+} and S_{an-} were adjusted in the simulations to tune the initial pH to fit the experiments. These parameters were adjusted because they only affect the pH, but otherwise can be treated as inerts in the simulations (Batstone et al., 2002a).

2.4.2 Stoichiometry of sugar degradation to VFA

One important step in the anaerobic digestion of sugars is the acidogenesis of monosaccharides, and there are different acidogenic metabolic pathways in anaerobic digesters (Batstone et al., 2002a; Zhou et al., 2018). Which pathway that dominates, depends on the microorganisms, the substrate and the conditions in the reactor. The different pathways result in different ratios of the resulting VFAs.

In ADM1, reactions 1-3 describes the degradation of the glucose monomer to acetic acid, propionic acid and butyric acid.

1. Acetate: $C_6H_{12}O_6 + 2\ H_2O \rightarrow 2\ CH_3COOH + 2\ CO_2 + 4\ H_2$ (1)
2. Acetate, propionate: $3\ C_6H_{12}O_6 \rightarrow 4\ CH_3CH_2COOH + 2\ CH_3COOH + 2\ CO_2 + 2\ H_2O$ (2)
3. Butyrate: $C_6H_{12}O_6 \rightarrow CH_3CH_2CH_2COOH + 2\ CO_2 + 2\ H_2$ (3)

According to Batstone et al. (2002a) the first is the most common reaction in AD, followed by the second and third. This is graded in ADM1 by equations 4-7

giving the yields of acetic acid ($f_{ac,su}$), butyric acid ($f_{bu,su}$) and propionic acid ($f_{pro,su}$).

$Acetate$: $f_{ac,su} = 0.67 \cdot \eta_{1,su} + 0.22 \cdot \eta_{2,su}$ (4)

$Propionate$: $f_{pro,su} = 0.78 \cdot \eta_{2,su}$ (5)

$Butyrate$: $f_{bu,su} = 0.83 \cdot (1 - \eta_{2,su} - \eta_{1,su})$ (6)

$\eta_{3,su} = 1 - \eta_{2,su} - \eta_{1,su}$ (7)

$\eta_{1,su}$, $\eta_{2,su}$ and $\eta_{3,su}$ are the fractions of sugars that degrades through the reactions 1-3.

The η-values were varied in simulations to study the effect of changing the fractions of the different VFAs (Table 6). This was done to alternate the dominating acidogenic pathway. In these simulations, the fraction of acetic acid is decreased, the fraction of butyric acid is increased and the fraction of propionic acid is mostly kept constant but at a lower level than in the original model. All the other parameters were kept constant.

Table 6. Yields of VFA used in the simulations

Name of curve	$\eta_{1,su}$	$\eta_{2,su}$	$\eta_{3,su}$	Yields of VFAs
(original) ADM1	0.495	0.345	0.16	$f_{ac,su} = 0.41$
				$f_{pro,su} = 0.27$
				$f_{bu,su} = 0.13$
No_1	0.395	0.245	0.36	$f_{ac,su} = 0.32$
				$f_{pro,su} = 0.19$
				$f_{bu,su} = 0.30$
No_2	0.295	0.245	0.46	$f_{ac,su} = 0.25$
				$f_{pro,su} = 0.19$
				$f_{bu,su} = 0.38$
No_3	0.195	0.245	0.56	$f_{ac,su} = 0.18$
				$f_{pro,su} = 0.19$
				$f_{bu,su} = 0.46$
No_4	0.095	0.245	0.66	$f_{ac,su} = 0.12$
				$f_{pro,su} = 0.19$
				$f_{bu,su} = 0.55$
No_5	0.045	0.195	0.76	$f_{ac,su} = 0.07$
				$f_{pro,su} = 0.15$
				$f_{bu,su} = 0.63$

2.4.3 Uptake rate of monosaccharides

The uptake of monosaccharides is described by Monod kinetics (8).

$uptake_{su} = km_{su} \cdot X_{su} \cdot \frac{S_{su}}{S_{su} + Ks_{su}} \cdot I_{pH_{bac}} \cdot I_{NH_{limit}}$ (8)

where km_{su} is the maximum uptake rate for monosaccharide degrading organisms, S_{su} is the concentration of monosaccharides, Ks_{su} is the half saturation constant for monosaccharide degradation, X_{su} is the monosaccharide degrading organisms, $I_{pH_{bac}}$ is the inhibition caused by pH on acidogens and acetogens and

$I_{NH_{limit}}$ is the function to limit growth due to lack of inorganic nitrogen.

The parameter km_{su} was evaluated in relation to the experiments performed with the hemicellulosic sugars. All the other parameters were kept constant throughout these simulations. The simulations were performed by changing the maximum uptake rate for the monosaccharide degrading organisms, km_{su}. A starting-value of 30 gCOD gCOD^{-1} d^{-1}, proposed by Batstone et al. (2002a), was used for the degradation of sugar, and the value was varied in the range of 15-50 gCOD gCOD^{-1} d^{-1}.

2.4.4 First order step

ADM1 is structured in several main steps including disintegration, hydrolysis, acidogenesis, acetogenesis and methanogenesis. The rate limiting step for dissolved substrates is most often methanogenesis, while it is hydrolysis for particulate substrates. Including a first order step is done by adding the substrate as carbohydrates that need to undergo extracellular hydrolysis to form simple monosaccharides that can be taken up by the bacteria. In ADM1, hydrolysis of carbohydrates X_{ch} is already implemented with a simple first order rate expression (9) producing monosaccharides, S_{su}.

$$\rho = k_{hyd\_ch} \cdot X_{ch} \qquad (9)$$

where ρ is the hydrolysis rate of solid substrates (kg COD solid substrate m^{-3} d^{-1}), X_{ch} is the particulate carbohydrate concentration (kg COD solid substrate m^{-3}) and k_{hyd_ch} is the kinetic parameter for hydrolysis (d^{-1}). The hydrolysis constants, k_{hyd_ch}, (here not necessary representing hydrolysis), were estimated individually for each substrate using the parameter estimation function in Aquasim.

3 Results and Discussion

3.1 Experimental results

The biogas production from the five hemicellulosic sugars (Figure 1) had some variance in both the biogas production rate and methane yield. The biogas production from glucose and mannose started within a few hours, whereas a short lag-phase was observed for xylose, galactose and arabinose. The biogas production rate from arabinose were noticeable slower compared to the other sugars. The methane yields calculated (Figure 1) were in the range of 0.74-0.9 (gCOD$_{CH4}$/gCOD$_{Substrate}$), where glucose reached the highest yield and arabinose the lowest yield.
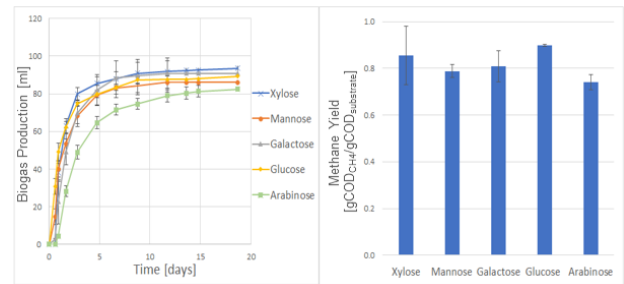
Figure 1. Experimental results from batch tests with hemicellulosic sugars. Left: Accumulated biogas production, Right: Methane yields [gCOD$_{CH4}$/gCOD$_{Substrate}$]

3.2 Simulations

The simulation results obtained with ADM1 in Aquasim were compared to the experimental results from the five hemicellulosic sugars.

3.2.1 Variation of the monosaccharide degradation stoichiometry

Changing the stoichiometry for monosaccharide degradation producing VFA had low impact within the variations tested.

At a load of 10 kgCOD m^{-3} used for the hemicellulosic sugars both in the experiments and the simulations, the VFA accumulation was low and the degradation of the VFAs fast, not influencing the biogas production rate or the pH (Figure 2). At higher load, this mechanism might be of higher importance due to possible VFA accumulation and the difference in the degradation rate of the VFAs (Batstone et al., 2002a).
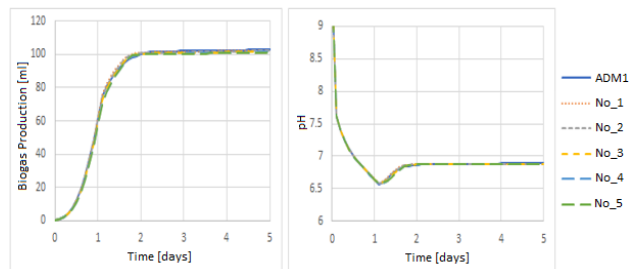
Figure 2. Simulation results with varying fractions of the VFAs (named ADM1 to No_5, Table 6). Left: Accumulated biogas production, Right: pH

3.2.2 Variations of the monosaccharide degradation kinetics

The simulation results indicate that the km-value has a clear impact on the initial biogas production (Figure 3). An increase in the km-value increases the uptake rate of monosaccharides, increase the concentration of VFAs and reduce the pH (Figure 3). Initially, the gas production is higher with the highest km-values because of the higher uptake rates of the organisms. The fit with experimental values, taking mannose as a model substrate, did not improve by changing the km-values

from 15 to 50. The total biogas produced reaches the same level independently of the km-value (rate).
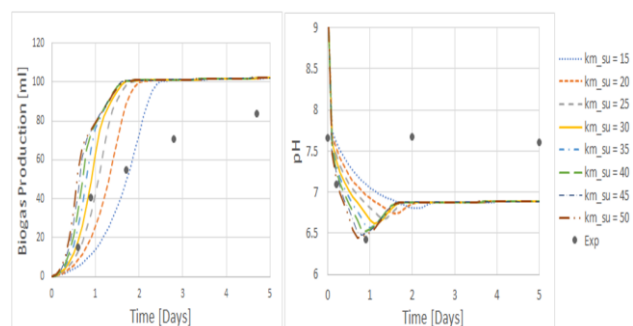
Figure 3. Simulation results with changes in the km-value. Experimental values are from mannose. Left: Accumulated biogas production, Right: pH

3.2.3 Simulations with addition of a first order step

The simulations adding the sugars as monosaccharides, without the first order step, gave only moderate fit with the experimental results (Figure 4) due to fast degradation of sCOD, and the experimentally observed lag-phases were not seen in the simulations, with the largest difference for arabinose.

Including the concentration dependent first order rate expression clearly improved the fit between simulated and experimental biogas production (Figure 4). The largest improvements were obtained for the sugars with an observed lag-phase during the experiments (arabinose, galactose and xylose). Glucose, the model monomer used in ADM1, obtained the lowest improvement of the five sugars, but still improved noticeably.
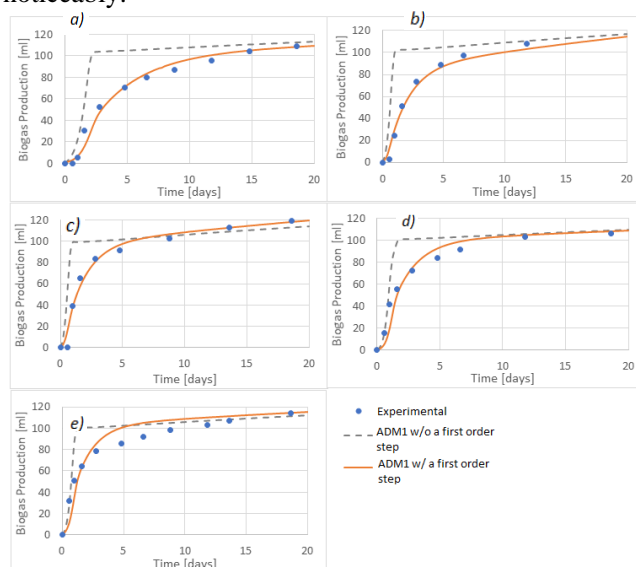
Figure 4. Simulated accumulated biogas production without a first order step (dashed line), with an added first order step (solid line) and experimental for a) Arabinose, b) Galactose, c) Xylose, d) Mannose and e) Glucose

The individual first order constants are estimated with parameter estimation in Aquasim (Table 7). The experimentally measured sCOD and the simulated substrate COD show the same trends in degradation, which indicates a resembling degradation rate (results not included here).

Table 7. Estimated first order constants

Sugar	Estimated first order constant $[d^{-1}]$
Glucose	0.64
Mannose	0.52
Xylose	0.61
Galactose	1
Arabinose	0.27

It is not obvious from the sugar chemistry that the carbohydrates used as substrate here need to undergo hydrolysis to be taken up by the acid forming bacteria. However, the simulation results clearly suggest that including the hydrolysis step in ADM1 in the form of a concentration dependent first order rate expression improves the simulations. Due to the simplicity and universality of this step, there are different processes related to sugar degradation that should be considered in this perspective. The uptake and degradation processes by the different monosaccharides could be more complex than included in ADM1. In addition, the types of microorganisms present could be important. Certain microorganisms mainly digest hexoses, while others digest both hexoses and pentoses (Chandrakant and Bisaria, 1998; Larsson, 1999). The ratio between these and the adaption period can be important regarding the sugar degradation rate. Hence, knowledge of the available microorganisms is important to modify ADM1 by distinguishing between the microorganisms as proposed by Ramirez et al. (2009).

Another consideration is the difference in degradation of sCOD in the reactors. Experimentally, the biogas production slows down as the sCOD in the batch reactor decreases. In the simulations, the biogas production continues at the same rate until the sugar is completely degraded. This difference might be caused by diffusion and lower accessibility of substrate for the microorganisms as the concentration of available substrate is lower. The simulations account for the decrease in substrate concentration with time, but the lower accessibility is not accounted for.

It is difficult to determine the exact process or the sum of processes that evidently makes it appropriate to include the first order rate expression. Of this reason, it is concluded that a step corresponding to the hydrolysis step in ADM1 could be included to model these reactions since it can be the rate limiting step. Further studies of the uptake and degradation kinetics of simple sugars are suggested to elucidate the underlying mechanisms.

4 Conclusion

The ADM1 model is a good tool to predict the biogas production from the hemicellulosic sugars during AD. The simulation results show moderate to good agreement with the experimental results depending on some key parameters. Adjusting the maximum substrate uptake rate of monosaccharides (km_{su}) and the stoichiometry during monosaccharide degradation to the VFAs acetic acid, propionic acid and butyric acid (by varying $\eta_{1,su}$, $\eta_{2,su}$ and $\eta_{3,su}$) had little influence on simulation fits to experimental results. The first order reaction step, usually accounting for hydrolysis, improved the fit between simulations and experimental biogas production from the hemicellulosic sugars, including the observed lag-phases. It is argued that the underlying mechanism, however, is not hydrolysis but rather related to type of microorganism and diffusion limitations.

Acknowledgements

References

T.E. Amidon and S. Liu. Water-based woody biorefinery. *Biotechnology Advances,* 27(5): 542-550, 2009. https://doi.org/10.1016/j.biotechadv.2009.04.012

B.E. Baeta, H.J. Luna, A.L. Sanson, S.Q. Silva, and S.F. Aquino. Degradation of a model azo dye in submerged anaerobic membrane bioreactor (SAMBR) operated with powdered activated carbon (PAC). *Journal of Environmental Management,* 128: 462-470, 2013. https://doi.org/10.1016/j.jenvman.2013.05.038

D.J. Batstone, J. Keller, I. Angelidaki, S.V. Kalyuzhny, S.G. Pavlostathis, A. Rozzi, W.T.M. Sanders, H. Siegrist, and V.A. Vavilin. *Anaerobic digestion model No. 1 (ADM1),* IWA Publishing, London, UK, 2002a.

D.J. Batstone, J. Keller, I. Angelidaki, S. Kalyuzhnyi, S. Pavlostathis, A. Rozzi, W. Sanders, H. Siegrist, and V. Vavilin. The IWA anaerobic digestion model no 1 (ADM1). *Water Science and Technology*, 45(10): 65-73, 2002b. https://doi.org/10.2166/wst.2002.0292

W.H. Bergland, C. Dinamarca, M. Toradzadegan, A.S. Nordgard, I. Bakke, and R. Bakke. High rate manure supernatant digestion. *Water Research*, 76, 1-9, 2015. https://doi.org/10.1016/j.watres.2015.02.051

P. Chandrakant and V.S. Bisaria. Simultaneous Bioconversion of Cellulose and Hemicellulose to ethanol. *Critical Reviews in Biotechnology*, 18(4): 295-331, 1998. https://doi.org/10.1080/0738-859891224185

T. Gehring, M. Lübken, K. Koch, and M. Wichern. ADM1 simulation of the thermophilic mono-fermentation of maize silage–Use of an uncertainty analysis for substrate characterization, *13th World Congress on Anaerobic Digestion*, 28, 2013.

F. Hu and A. Ragauskas. Pretreatment and Lignocellulosic Chemistry. *BioEnergy Research*, 5(4): 1043-1066, 2012. https://doi.org/10.1007/s12155-012-9208-0

S. Larsson, E. Palmqvist, B. Hahn-Hägerdahl, C. Tengborg, K. Stenberg, G. Zacchi, and N.-O. Nilvebrant. The generation of fermentation inhibitors during dilute acid hydrolysis of softwood. *Enzyme and Microbial Technology,* 24: 151-159, 1999. https://doi.org/10.1016/S0141-0229(98)00101-X

J.E. McMurry. *Organic Chemistry with Biological Applications*. Cengage Learning, 2014.

K. Ostgaard, V. Kowarz, W. Shuai, I.A. Henry, M. Sposob, H.H. Haugen, and R. Bakke. Syringe test screening of microbial gas production activity: Cases denitrification and biogas formation. *Journal of Microbiology Methods*, 132: 119-124, 2017. https://doi.org/10.1016/j.mimet.2016.11.021

E. Palmqvist and B. Hahn-Haagerdal. Fermentation of lignocellulosic hydrolysates II inhibitors and mechanisms of inhibition. *Bioresource Technology*, 74: 25-33, 2000. https://doi.org/10.1016/S0960-8524(99)00161-3

I. Ramirez, E.I.P. Volcke, R. Rajinikanth, and J.P. Steyer. Modelling microbial diversity in anaerobic digestion through extended ADM1 model. *Water Research,* 43: 2787-2800, 2009. https://doi.org/10.1016/j.watres.2009.03.034

M. Zhou, B. Yan, J.W.C. Wong, J. W. C. and Y. Zhang. Enhanced volatile fatty acids production from anaerobic fermentation of food waste: A mini-review focusing on acidogenic metabolic pathways. *Bioresource Technology*, 248: 68-78, 2018. https://doi.org/10.1016/j.biortech.2017.06.121

Simulation of a purification plant for pyrolysis gas based on plastic waste

Ebrahim Qaredaghi [1]   Sara Ronasi[2]   Britt M. E. Moldestad[1]

[1] Department of Process, Energy and Environmental Technology, University of South-Eastern Norway, Norway,
qaredaghi@gmail.com Britt.Moldestad@usn.no
[2]Norner Research, Norway, sara.ronasi@norner.no

Abstract

This study focuses on exctracting ethylene and propylene from other components in the pyrolysis gas mixture of plastic waste. Selective absorption of ethylene and propylene from the gas mixture by using a silver nitrate solution, was selected as a promising technology. A lab scale set-up was built and experimental tests were performed using a model gas mixture. Aspen HYSYSV10 was used to model and simulate the absorption process. The model was validated against experimental data. The validated model was further used to identify improvements for the separation process and increase the recovery of ethylene and propylene. The results from the simulated improved process show that the amount of ethylene and propylene in the product gas could be significantly increased. In the experimental study, only 25% of ethylene and propylene in the feed was captured, whereas the simulation of the improved process indicates that the recovery of the monomers could be almost quantitative (99%). However, the product gas from the separation system contained an undesirable high amount of CO and CO_2. These gases act as pollutants of the polyolefin reaction and further studies are needed to obtain pure olefin gases.

Keywords: plastic waste, purification, ethylene, propylene, absorption, pollutants, Aspen Hysys, process simulation

1   Introduction

Global production of plastics has since 1940 become one of the fastest growing industries in the world. PlasticsEurope (PlasticsEurope, 2019) and the European Association of Plastics Recycling and Recovery Organizations (EPRO, 2018) reported that the global plastic production in 2017 was 348 million tons, out of which 64.4 million tons were produced in Europe. The largest plastic producers were China and Europe by producing 29.4% and 18.5% of the total respectively. (G. Gourmelon, 2015; A. Demirbas, 2004) The European commission has presented a plan to implement a circular economy with a key focus on plastics. The plan is set to ensure that all plastic packaging is reused or recycled by the year 2030.

(Plastics Europe, 2019). According to EUs circular Economy Package and directive (EU) 2018/850 adopted by the European parliament and by the EU council in 2018, Member States will be required to ensure that, as of 2030, waste suitable for recycling or other recovery, in particular contained in municipal waste, will not be permitted to be disposed of to landfill. The re-processing operation in mechanical recycling is not cost effective since this process has high energy demand for sorting, cleaning, transportation and processing (Panda et al., 2010). Chemical recycling is the state of the art feedstock recycling, also known as tertiary recycling. The aim is to convert waste polymers into valuable chemicals or original monomers. Cracking is one of the main approaches to this method (Panda et al., 2010).

The aim of this study is a) to carry out experimental separation tests to obtain pure ethylene and propylene, and b) to develop a process simulation model that can be used to optimize the purification process. In order to obtain the requested purity grade for polymerization of ethylene and propylene, there are two options. The first option is to focus on ethylene and propylene and try to extract them from the other components in the pyrolysis gas. The second option is to remove all impurities in the pyrolysis gas one by one. The first option was chosen in this study, and absorption is used to extract ethylene and propylene from a gas mixture.

2   Purification process

Figure 1 illustrates the general steps of sustainable polymer production from biomass or plastic waste. Plastic waste or biomass is fed to the pyrolysis reactor. The composition of the outlet gas from the reactor is highly dependent on the composition of the plastic waste. In the pyrolysis process, the components of the plastics can be converted into solid (char), liquid (pyrolysis oil) and gas (Dermibas, 2004). The product gas is collected from the top of the pyrolysis reactor, and in order to use the gas in the plastic production process, a purification process is needed. The polymerization reaction is sensitive to impurities, and all impurities should be removed from the gas. Therefore, the purification process has an essential role in the whole process. The products from the polymerization process are polyethylene (PE) and polypropylene (PP).
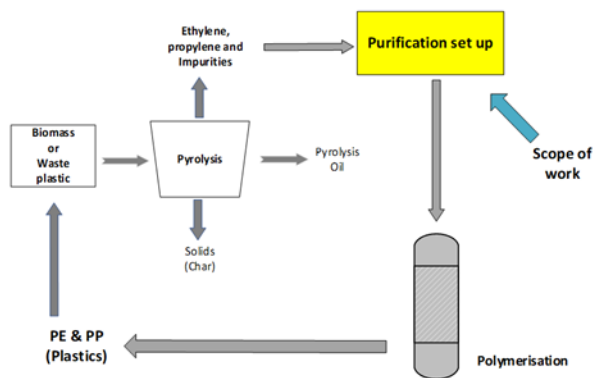
Figure 1. Simplified overview of the plastic recycling process.

Thermal pyrolysis is a well-known method for managing and recycling plastic waste. The composition of the product gases from the pyrolysis reactor can vary due to differences in the composition of the plastic wastes used as feed to the pyrolysis reactor. Hydrogen, oxygen, light olefins (ethylene, propylene, butene), aromatics (benzene, toluene, xylene), polycyclic aromatic hydrocarbons (PAHs), carbon monoxide (CO), carbon dioxide (CO_2) are the main gaseous component (Donaj et al., 2012; Xue et al., 2016).

Several methods and technologies to be used for purification of the pyrolysis gases are found in literature. An absorption process based on the silver-olefin complexation mechanism was chosen to extract ethylene and propylene from the other gas components (Ananthapadmanabhan and Goddard, 1988; Safarik and Eldridge, 1998; Li, 2013). The advantages of the process is that it can selectively extract the ethylene and propylene from the pyrolysis gas as the silver ions bind ethylene and propylene while they do not interact with CO and CO_2. CO and CO_2 are considered as poisons for the polymerization process, and it is crucial to minimize the content of those components in the purified gas. A disadvantage with the process is that the silver nitrate is expensive, and it is very important that it can be regenerated without significant degradation.

Mortaheb *et al* developed a mathematical model and studied the absorption of ethylene from an ethylene-ethane gas mixture by using a silver nitrate solution in a semi-continuous process. The model includes that absorption is a function of temperature and concentration of the absorbing solution. According to the experimental results, an increase in temperature decreases the amount of absorbed ethylene. Furthermore, the amount of absorbed ethylene is increased in solutions with higher concentrations of silver nitrate. Figure 2 compares the total absorption values from the model with experimental data for two different silver nitrate molarities (Mortaheb et al., 2009).
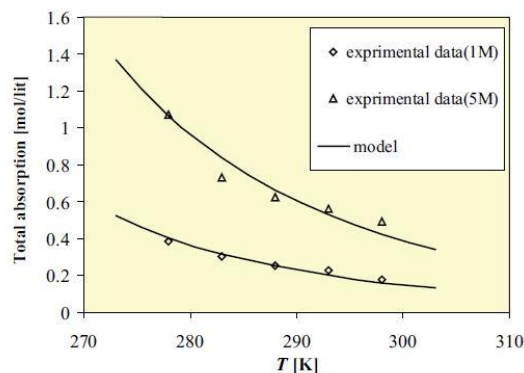
Figure 2. Comparison of total absorption with estimated values by the model (Mortaheb *et al*, 2009).

A mathematical model proposed by Ghasem *et al* presents the solubility of ethylene as a function of silver nitrate concentration, feed gas pressure and temperature. Ghasem *et al* studied the separation of ethylene from an ethylene/ethane gas mixture using a hollow fiber membrane. The model is valid at pressure up to 0.64 MPa, silver nitrate concentrations from 1 to 6 M and temperatures from 278 to 308 K. Figure 3 shows the ethylene solubility in silver nitrate solution as a function of the feed gas pressure and the silver nitrate concentration. The solubility of ethylene increases with increasing pressure and with increasing concentration of the silver nitrate solution (Ghasem *et al*, 2017).
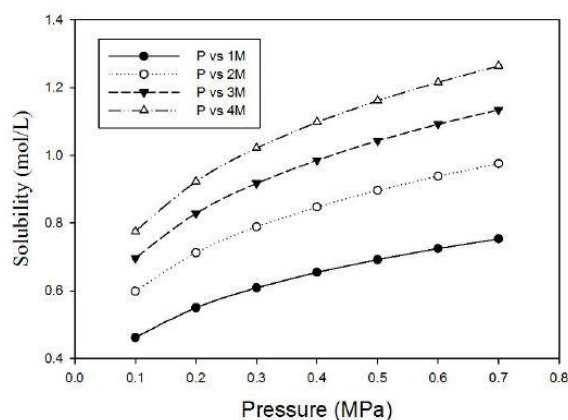
Figure 3. Ethylene solubility in silver nitrate solution as a function of gas pressure and silver nitrate concentration (Ghasem *et al*, 2017).

3 Material and methods

3.1 Experimental set-up

The experimental set-up is shown in Figure 4. The test rig consists of a reactor where the aqueous silver nitrate solution is filled. The reactor is connected to a high-pressure vessel which is again connected to the main gas mixture capsule. The piping from the top of the reactor is connected to vent, to an aluminum gas collection bag and to a vacuum bomb. Each line has valves to control the flow. A digital temperature indicator and a pressure

indicator are connected to the reactor to monitor the temperature and pressure inside the reactor. Nitrogen is connected to the feeding line, and is used to purge the system. A gas chromatograph (GC), with helium as the carrier gas, was used for the analysis of the gas samples.
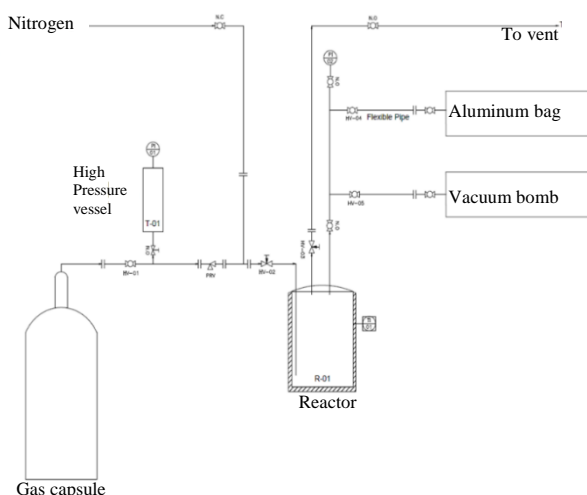
Figure 4. Experimental set-up for extraction of ethylene and propylene.

3.2 Simulation model

The process simulation software Aspen HYSYSV10 is used to model and simulate the purification process. There are several thermodynamic models available in the Aspen HYSYSV10 library. In order to obtain accurate simulation results, it is crucial to choose the appropriate thermodynamic model. The three main thermodynamic groups of models are the equation of state (EOS), the activity coefficient models and some special models which are related to specific components. In this study an activity coefficient model was used. Activity coefficient models are applied for highly non-ideal polar systems, and are empirical models. In these models, an equation of state is used for predicting the vapor fugacity coefficients and an activity coefficient model is used for the liquid phase (Aspen HYSYS Software, 2014). Different models were tested, and the Electrolyte Non-Random Two-Liquid (eNRTL) model was chosen for the simulations. The reason for choosing the eNRTL model, is that silver nitrate is a strong electrolyte (Keller *et al*, 1992). The eNRTL model covers a wide variety of aqueous and mixed-solvent electrolyte systems covering the whole concentration range from fused salts or pure solvents to saturated solutions (Chen and Song, 2004).

Figure 5 shows a set-up for the absorption process developed in Aspen HYSYS. The absorbent silver nitrate (Lean Solution) and the gas mixture (Feed Gas) are fed to the absorption tower (Absorber Tower-100) to extract ethylene and propylene from the gas mixture. The pressure in the absorber is the same as in the feed gas and the absorbent. The absorption is the first step in the ethylene and propylene separation.

The rich silver nitrate solution from the absorber contains the absorbed ethylene and propylene together with some impurities. The solution is fed to a distillation tower (Distillation Tower-102), where most of the impurities goes to the top of the tower and to vent. The absorbent including ethylene and propylene is taken out in the bottom, and fed to the next distillation tower (Distillation Tower-103), which operates at a lower pressure. In this tower ethylene and propylene are separated from the silver nitrate solution. The distillation tower (Distillation Tower-103) is operated under vacuum to capture as much as possible of the ethylene and propylene gases from the solution. The regenerated lean absorbent is taken out in the bottom of the distillation tower, and is recycled to the absorption tower. In order to ensure that the temperature and pressure of the recycled absorbent is the same as in the absorption tower, the Set-functions are defined to adjust the pressure after the pump (Pump-100) and the temperature after the heat exchanger (Heat-Exchanger-100). The recycle function (RCY-1) in the regeneration loop, is defined to update the estimated values in the lean absorbent stream.

Since the extraction of ethylene and propylene with silver nitrate follows the π-complexation mechanism and does not include chemical reactions, distillation columns are used for the separation of the components from the absorber. In order to develop a robust model to simulate the ethylene and propylene absorption, the model has to be validated against the experimental data. A one-stage absorber, as used in the experimental study, is therefore chosen for the development of the model. The validated model can further be improved by including multiple stages, in order to obtain the desired separation. A packed absorber is used in the simulations. The reason for choosing a packed absorber instead of the tray type, is to minimize the inventory of silver nitrate in the absorber (Keller *et al*, 1992). The pressure in the absorber is 1700 kPa. The diameter of the absorber is 1.5 m and the packing height is 0.61 m.
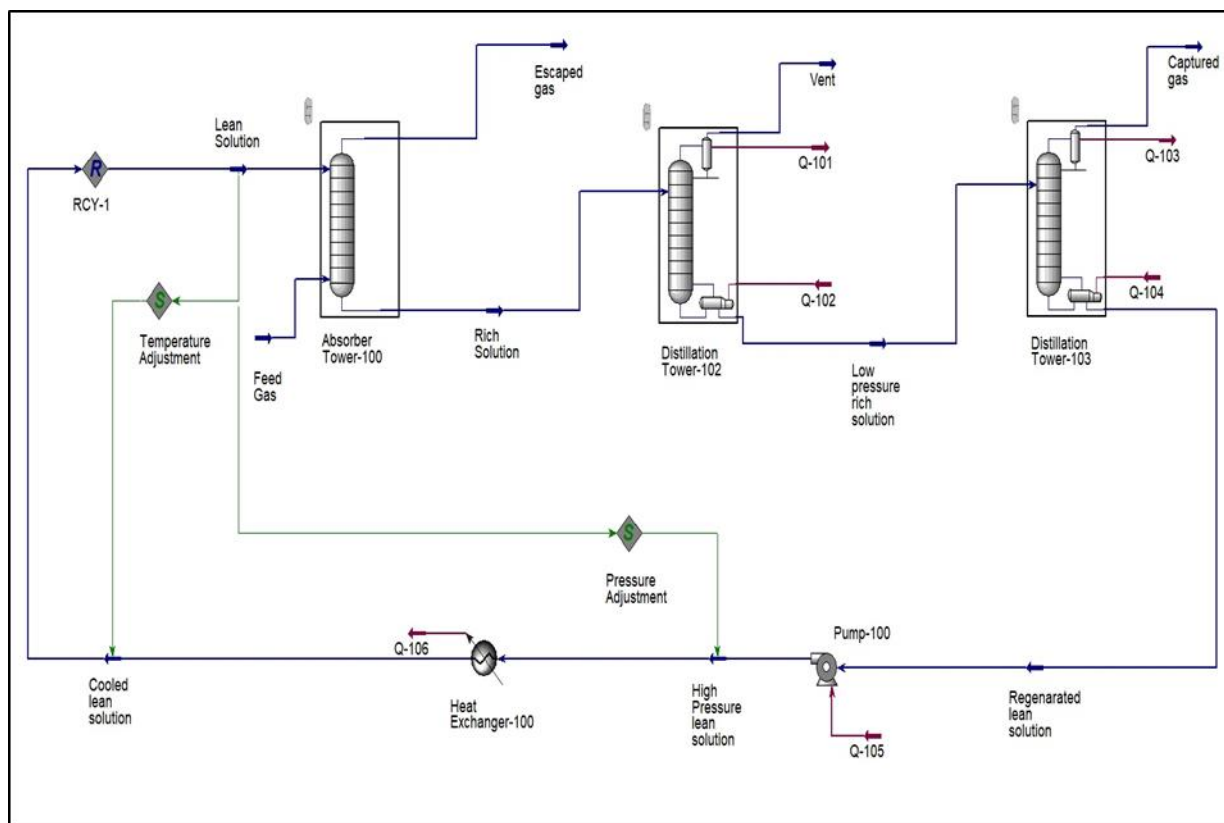
Figure 5. Simulation model for ethylene and propylene separation

4 Results

4.1 Experimental results

The solubility of ethylene and propylene in a silver nitrate solution is a function of gas pressure and silver nitrate concentration. Experiments were carried out using a 3M silver nitrate solution at ambient temperature and at initial pressure of 41 bar. The composition of the feed gas to the absorber is presented in Table 1.

Table 1. Feed gas composition.

Component name	Mass fraction	Mole fraction
C_2H_4	0.2	0.2185
C_3H_6	0.2	0.1456
CO	0.05	0.0546
CO_2	0.05	0.0349
N_2	0.5	0.5463

A series of experiments were carried out to study the absorption of ethylene and propylene in silver nitrate. Two of the experiments are presented in this paper. The pressures in the different tanks are presented in Table 2. The same pressures were used as input to the simulation model. One of the big challenges during the experiments, was to obtain exactly the same pressure in all the tests. In theory, the absorption of olefins increases with increasing pressure (Keller *et al,* 1992). In the

experimental setup, it was very difficult to maintain a stable pressure during the absorption at pressures higher than 17 bar (final pressure R1). The pressure was therefore kept at 17 bar, but still some pressure variations between the experiments were observed. Therefore, only two experiments (one with new silver nitrate solution and one with reused solution) are presented here. The results from these two experiments will further be used for validation of the process simulation model. The pressure in the vacuum bomb and the aluminum bag are critical for the results of the tests.

Table 2. Operating pressures for the absorption process

Pressure [bara]	Exp. 1 (New solution)	Exp. 2 (Reused solution)
Initial pressure (R-01)	41.03	41.01
Final pressure (R-01)	17.01	17.01
Vacuum bomb	5.51	5.51
Initial alum. bag	17.01	17.01
Final alum. bag	1.05	1.05

Table 3 presents the results of the experiments (Kannan, 2019), and shows the recovery of the different gas components in the captured gas. Due to the limitations in the analysis set-up, nitrogen was injected into the vacuum bomb to enable feeding of the sample to the gas chromatograph.

Table 3. Mole % recovery of each component in the captured gas.

	CO_2	CO	N_2	C_2H_4	C_3H_6
Ex. 1	8.4	1.82	0.01	25.5	25.5
Ex. 2	7.5	1.26	0.01	25.6	24.7

Based on the results and the mass balance over the different units in the set-up, the mass flow rate for each stream was calculated (Kannan, 2019). The results from the experiments were used to design a cleaning process for an inlet gas flow rate of 10 kg/h. Table 4 presents the calculated mass flow rates for the different streams based on the two experiments.

Table 4. Mass flow rates for the process streams.

	Exp. 1 [kg/h]	Exp. 2 [kg/h]
Inlet silver nitrate solution [kg/h]	189	203
Inlet gas flow [kg/h]	10.0	10.0
Escaped gas [kg/h]	7.70	7.70
Vent gas [kg/h]	0.85	0.91
Captured gas [kg/h]	1.40	1.40

As can be seen from Table 4, a large part of the feed gas leaves the absorber as escaped gas, which suggests need for improvement of the experimental set-up.

4.2 Simulation results

The process simulation is validated against the experimental results, and the validated model is further improved to get a higher recovery of ethylene and propylene in the captured gas.

4.2.1 Validation of model

Table 5 shows the simulated and the experimental flow rates of the captured and escaped gas. The simulated flow rates of the captured gas are higher than the experimental results, whereas the escaped gas flow rates are higher in the experiments than in the simulations. This can be due to more ideal and stable operation conditions in the simulations.

Figure 6 shows a comparison of the simulated and experimental composition of the captured gas. Some deviations between the computational and experimental results are observed. The experimental mole fraction of ethyelene are significantly higher than the simulated mole fraction, whereas the experimental mole fractions for propylene, CO_2 and CO are slightly lower than in the simulations. The reason may be variations in the feed of silver nitrate to the absorber. The deviations can also be explained by the tray efficiency, which had to be adjusted in the software in order to the obtain convergence.

In Table 6, the recovery of the different components in the captured gas is presented. The recovery of ethylene and propylene is significantly higher than in the experiments, which is beneficial for the extraction process. However, the process simulation also gives a significantly higher absorption of the unwanted components CO and CO_2.

Table 5. Mass flow rates for the process streams.

	Flow rate capured gas [kg/h]	Flow rate escaped gas [kg/h]
Exp. 1	1.4	7.7
Exp. 2	1.4	7.7
Sim. 1	1.8	5.8
Sim. 2	1.7	5.8
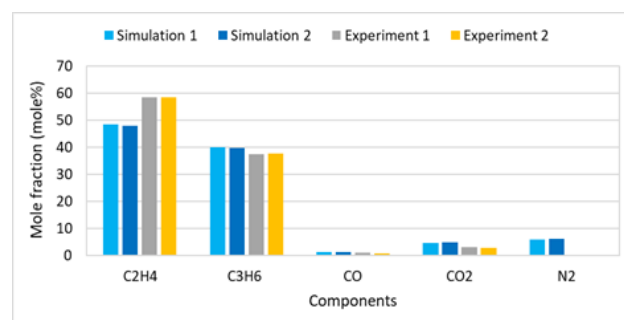
Figure 6. Comparison of simulated and experimental composition of captured gas.

Table 6. Simulated mole % recovery of each component in the captured gas.

	CO_2	CO	N_2	C_2H_4	C_3H_6
Sim. 1	19.9	3.32	0.01	33.9	42.0
Sim. 2	22.0	3.92	0.01	34.2	41.2

Figure 7 shows the comparison of the composition of the escaped gas. There are only small deviations between the experiments and the simulations. The escaped gas contains about 60 mole% nitrogen, but it also contains about 20% ethylene and 12% propylene in addition to the CO and CO_2. This means that there is a large potential to recover more of the ethylene and propylene.
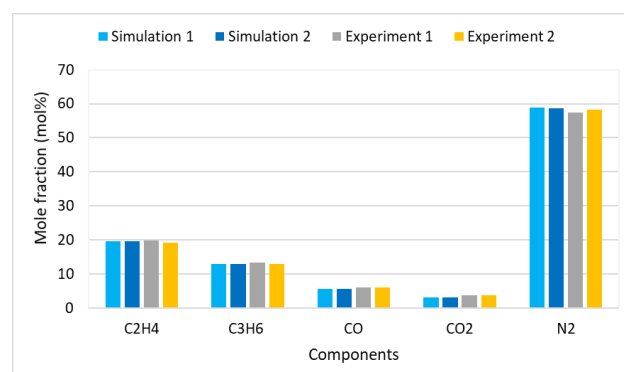
Figure 7. Comparison of simulated and experimental composition of escaped gas.

The simulated loss of each component through the escaped gas are presented in Table 7. The loss is defined

as the mole % of the components in the feed leaving the absorber in the outlet gas. As can be seen from the table, more than half of the ethylene and propylene are lost in the outlet gas stream from the absorber.

Table 7. Simulated loss (in mole % of the feed) of the components through the escaped gas.

	CO_2	CO	N_2	C_2H_4	C_3H_6
Sim. 1	50.7	60.8	63.4	52.8	52.2
Sim. 2	50.6	60.7	63.2	52.9	52.3

4.2.2 Improvement of the simulation model

In order to absorb all the ethyelene and propylene using the silver nitrate solution in the absorber, it is necessary to change parameters that affect the escaped gas quantity and quality. It is important here to look at the number of trays and the tray efficiency in the absorber. The number of trays are therefore increased from 1 to 5, and the tray efficiency are increased from 0.5 to 0.7. The actual tray efficiency depends on the flow rate of the feed gas and the absorbent. By increasing both these parameters in the absorber, almost all ethylene and propylene are solved in the silver nitrate solution. Packed absorbers are most efficient and have been used both for the validation and the improvement of the model.

The separation in the first distillation tower (Tower-102), depends very much on the pressure. By specifying the pressure in the first distillation tower to 750 kPa, a significant amount of N_2 and a small amount of CO is leaving in the vent gas, and the bottom product of the tower consists mainly of the silver nitrate solution, ethyelene and propylene in addition to some CO and CO_2. The second distillation tower (Tower-103) can be operated at the same pressure (61.3 kPa) as in the base case. The number of trays is kept at 3 in the first distillation tower and are increased from 1 to 2 in the second tower. The tray efficiency in both the distillation towers is 0.6, and the reflux ratio is set to 0.3. The simulation of the improved model was run with a mass flow rate of 234 kg/h for the lean silver nitrate solution. The improved model gave a mass flow rate of 2.54 kg/h for the outlet gas from the absorber, 2.2 kg/h for the vent gas from the second distillation tower and 5.27 kg/h for the captured gas. This is a significant improvement compared to the experimental results and the results from the base case simulations.

Figure 8 shows the gas composition in the captured gas. About 2 kg of ethylene and 2 kg/h of propylene are recovered, which means a recovery of nearly 100%. However, at the same time, the recovery of CO, CO_2 and nitrogen has also increased, which is not beneficial. A further improvement of the process is therefore needed as CO and CO_2 are considered as poisons for the polymerization process, and it is crucial to minimize the content of those components in the purified gas. One solution is to change the temperature and pressure both in the absorber and the distillation towers. Increasing the temperature in the absorber, may reduce the absorption

of CO and CO_2, but it will also reduce the recovery of ethylene and propylene. Reducing the pressure in the distillation towers will release more of captured CO, CO_2 and N_2 to the vent gas. The content of CO and CO_2 has to be reduced to very low values, and a recovery of ethyelene and propylene in the range of 50-70% may therefore be acceptable.
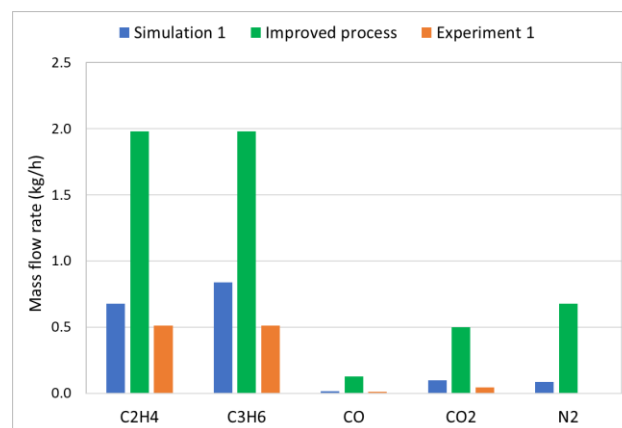
Figure 8. Comparison of the composition in the captured gas.

5 Conclusion

In order to extract ethylene and propylene from the plastic waste pyrolysis process, absorption process using silver nitrate solution as absorbent was selected as a promising technic. A lab scale set-up was built and experimental tests were performed to study the suitability of silver nitrate solution for extracting the hydrocarbons from the pollutants.

Aspen HYSYSV10 with the property package Electrolyte NRTL was used to model and simulate the absorption process. The model was validated against experimental data. The results from the simulations agreed well with the experimental data in terms of the mole fractions and the flow rates of captured gas. The validated model was further used to improve the separation process and increase the recovery of ethylene and propylene. The results from the improved process show that the amount of ethylene and propylene in the product gas increased significantly. In the experimental study, only 25% of ethylene and propylene in the feed was captured, whereas the simulation of the improved process increased the recovery of the monomers to 99%. The product gas contains a relatively high amount of CO and CO_2. The catalytic polymerization process will be poisoned or deactivated if CO and CO_2 are present in the monomer feed gas, and it is crucial to reduce the concentration of these components below acceptable levels. The process simulation model has to be improved to obtain the low limits for CO_2 and CO.

Acknowledgements

References

K. P. Ananthapadmanabhan and E. D. Goddard. Process for the recovery of metal values, United States, 1988. Available: http://www.freepatentsonline.com/4765834.html

Aspen. HYSYS, Software, and V8.6, Aspen HYSYS help, 2014.

C. C. Chen and Y. Song. Generalized electrolyte-NRTL model for mixed-solvent electrolyte systems, *AIChE Journal,* 50(8): 1928-1941, 2004.

A. Demirbas. Pyrolysis of municipal plastic wastes for recovery of gasoline-range hydrocarbons, *Journal of Analytical and Applied Pyrolysis,* 72(1): 97-102, 2004.

P. J. Donaj, W. Kaminsky, F. Buzeto, and W. Yang. Pyrolysis of polyolefins for increasing the yield of monomers' recovery, *Waste management,* 32(5): 840-846, 2012.

EPRO. *Best Recycled Plastic Product Award*, 2020, http://www.epro-plasticsrecycling.org/

N. Ghasem, M. Al-Marzouqi, and N. Sheta. Effect of Pressure on the Separation of Ethylene from Ethylene/Ethane Gas Mixture Using Hollow Fiber Membrane, *Am. J. Polym. Sci,* 7: 30-37, 2017.

G. Gourmelon. Global plastic production rises, recycling lags, *Worldwatch Institute analysis explores trends in global plastic consumption and recycling,* 2015.

D. K. Kannan. Process design of lab scale purification tool, Master thesis, University of South-Eastern Norway, 2019.

G. E. Keller, A. E. Marcinkowsky, S. K. Verma, and K. D. Williamson. Olefin recovery and purification via silver complexation, *Separation and Purification Technology,* 1, 1992.

N. Li. Separation and Purification Technology. CRC Press, 2013

H. Mortaheb, M. Mafi, A. Zolfaghari, B. Mokhtarani, N. Khodapanah, and F. Ghaemmaghami. Absorption of ethylene from ethylene-ethane gaseous mixture by AgNO3 solution in a semi-continuous process, *The Canadian Journal of Chemical Engineering,* 87(6): 957-964, 2009.

A. K. Panda, R. K. Singh, and D. K. Mishra. Thermolysis of waste plastics to liquid fuel: A suitable method for plastic waste management and manufacture of value added products—A world prospective, *Renewable and Sustainable Energy Reviews,* 14(1): 233-248, 2010.

PlasticsEurope. *Plastics - the Facts 2018*, https://www.plasticseurope.org/application/files/6315/4510/9658/Plastics_the_facts_2018_AF_web.pdf

D. J. Safarik and R. B. Eldridge. Olefin/Paraffin Separations by Reactive Absorption: A Review, *Industrial & Engineering Chemistry Research*, 37(7): 2571-2581, 1998.

Y. Xue, A. Kelkar, and X. Bai. Catalytic co-pyrolysis of biomass and polyethylene in a tandem micropyrolyzer, *Fuel,* 166: 227-236, 2016.

Simulation of condensation in raw biogas containing H₂S

Terje Bråthen[1] Lars Erik Øi[1] Jon Hovland[2]

[1]Department of and Process, Energy and Environmental Technology, University of South-Eastern Norway
[2]SINTEF Tel-Tek, SINTEF Industry, Porsgrunn, Norway

`lars.oi@usn.no`

Abstract

Condensation in raw biogas during compression is a problem because the CO_2, water and H_2S in the liquid phase is very corrosive. Raw biogas typically contains 60 mol-% methane, 40 mol-% CO_2, is saturated with water and may contain contaminants as H_2S. In case of H_2S, it is of interest whether it has influence on the dew point (condensation) temperature. The aim of this work is to calculate the dew point under different conditions using different equilibrium models. Phase envelopes showing the two-phase area are also calculated. There is traditionally only one constant binary parameter for each component pair in standard models like Peng-Robinson (PR) and Soave-Redlich-Kwong (SRK). In the process simulation tools Aspen HYSYS and Aspen Plus, the binary parameter for water/CO_2 and water/H_2S can be made temperature dependent. For dry mixtures of methane and CO_2 with up to 1 mol-% H_2S (a high value for biogas), the different models gave similar results. When the H_2S increased from 0 to 1 mol-%, the dew point temperature increased with typically 1.0 K. When water was included, the amount of calculated H_2S dissolved in water varied considerably with the model and on the selected binary parameter for water/H_2S. For biogas simulation, it is recommended to select a binary parameter (fitted to -0.036 at 25 ℃ for PR) that fits the experimental data for H_2S solubility in water.

Keywords: CO₂, methane, water, biogas, phase envelope, Aspen HYSYS, Aspen Plus

1 Introduction

Condensation in biogas containing water and H_2S during compression is a challenge because the CO_2, water and H_2S in the liquid phase is very corrosive. Raw biogas typically contains 60 mol-% methane, 40 mol-% CO_2, is saturated with water and may contain contaminants as H_2S. Under compression up to 300 bar, it is a question whether condensation (mostly water) will occur. When the biogas contains H_2S, it is of interest whether it has influence on the dew point temperature. It is also of interest how much H_2S will condense in the liquid phase.

This work is a continuation of the work from Øi and Hovland (2018) and Bråthen et al. (2019) which did not consider H_2S.

Traditionally, gas mixture properties of methane, CO_2, H_2S and water are calculated in a process simulation program with standard models like PR (Peng and Robinson, 1976) and SRK (Soave, 1972). There is traditionally only one constant binary parameter for each component pair. In the process simulation tools Aspen HYSYS and Aspen Plus, the binary parameter for water/CO_2 and water/H_2S can be made temperature dependent. Other models are also available in Aspen HYSYS and Aspen Plus. Properties of mixtures of the biogas components have been studied extensively in natural gas processing where the methane concentration is very high. The same components have also been studied in the development of CO_2 injection into different hydrocarbon reservoirs (Ziabakhsh-Ganji and Kooi, 2012). There have been found few articles about process simulation of biogas (Tan et al., 2017; Pellegrini et al., 2015; Ahmad et al., 2018), but these are not considering H_2S.

When using fitted binary parameters (k_{ij} parameters) these models simulate the gas phase and the condensation point reasonably accurately (within a few degrees) at least below the critical point which is 46 bar for methane and 74 bar for CO_2 (Øi and Hovland 2018; Bråthen et al., 2019).

Studies of models for vapour/liquid equilibrium in the methane/CO_2/water-system have been performed by Jarne (2004) Austegard et al. (2006), Privat and Jaubert (2014), Al Ghafri et al. (2014), Legoix et al. (2017) and Bråthen et al. (2019). Only Privat and Jaubert (2014) included H_2S in the study. Austegard et al. (2006) concluded that a simple equation of state like SRK is satisfactory to describe the vapour phase, but more complex models are necessary to describe the liquid phase.

Several authors have studied models for the system CO_2/water (Spycher et al., 2003; Longhi, 2005; Aasen et al., 2017). Bråthen et al. (2019) concluded that PR and SRK with a standard k_{ij} value were able to simulate this system satisfactory, at least for the vapour phase up to the critical point (40-70 bar).

References for CH_4/CO_2/water including H_2S are Søreide and Whitson (1992) and Li et al. (2014). Søreide and Whitson calculated k_{ij} values fitted to

measured vapour phase compositions and different values fitted to liquid phase compositions.

There has been published work including H_2S using more advanced thermodynamic models in e.g. predictive PR (Privat et al., 2008), SAFT-VR (Perez et al., 2017). In this work, the emphasis is on the standard PR and SRK methods available in the process simulation programs Aspen HYSYS and Aspen Plus.

The aim of this work is to calculate the dew point or condensation of biogas containing H_2S under different conditions with varied temperature, pressure and gas composition and using different equilibrium models.

2   Simulation Programs and Models

Øi and Hovland (2018) and Bråthen et al. (2019) simulated dry biogas (CH_4 and CO_2) and mixtures also containing water using the process simulation programs Aspen HYSYS and Aspen Plus. The equilibrium models SRK (Soave, 1972), PR (Peng and Robinson, 1976) and also other models were used.

The reason why the simple models PR and SRK are selected, is that the models and fitted binary parameters are usually available in different process simulation programs. Other commercial process simulation programs are ProMax, ChemCad, and ProTreat.

In Aspen HYSYS and Aspen Plus the PR and SRK models have only one adjustable parameter for each binary component pair, but for some components, and especially water/CO_2 and water/H_2S, this parameter may be temperature dependent.

Equations for the SRK equation of state are shown in (1) to (8) from Aspen HYSYS Version 10. Aspen HYSYS and Aspen Plus Version 10 were used in the simulations. More details are discussed in Bråthen et al. (2019).

$$p = \frac{RT}{v-b} - \frac{a}{v(v+b)} \tag{1}$$

$$b = \sum_{i=1}^{N} x_i b_i \tag{2}$$

$$b_i = \frac{0,08664 R T_C}{p_c} \tag{3}$$

$$a = \sum_{i=1}^{N} \sum_{j=1}^{N} x_i x_j \left(a_i a_j\right)^{0,5} \left(1 - k_{ij}\right) \tag{4}$$

$$a_i = a_{ci} \alpha_i \tag{5}$$

$$a_{ci} = \frac{0,42748 R^2 T_c^2}{p_c} \tag{6}$$

$$\alpha_i = \left[1 + m_i \left(1 - T_r^{1/2}\right)\right]^2 \tag{7}$$

$$m_i = 0,48 + 1,574 \omega_i - 0,176 \omega_i^2 \tag{8}$$

P, T, v and R are the pressure, temperature, molar volume and the universal gas constant. T_c is the critical temperature, ω is the acentric factor and T_r is the reduced temperature (the ratio between T and T_c). The binary interaction parameter k_{ij} ($= k_{ji}$) is a constant for a binary component pair and x_i is the mole fraction for component i. In the PR equation, equation 1, 3, 6 and 8 are replaced by equation 9, 10, 11 and 12.

$$p = \frac{RT}{v-b} - \frac{a}{v(v+b)+b(v-b)} \tag{9}$$

$$b_i = \frac{0,077796 R T_C}{p_c} \tag{10}$$

$$a_{ci} = \frac{0,457235 R^2 T_c^2}{p_c} \tag{11}$$

$$m_i = 0,37464 + 1,54226 \omega_i - 0,26992 \omega_i^2 \tag{12}$$

The binary parameter k_{ij} is normally a constant for each binary pair. When utilizing the default k_{ij} values in Aspen HYSYS and Aspen Plus, the k_{ij} values are constant for all component pairs except for water/CO_2 and water/H_2S where it is a temperature dependent function. Different optimized values for the k_{ij} values can be found because the parameters may be optimized for different conditions, e.g., for accurate prediction of either the gas phase or the condensate phase. For the calculation of dew points, it is reasonable to use binary interaction coefficients optimized for the gas phase (Bråthen et al., 2019).

The PR and SRK versions used in Aspen Plus are equal to the Aspen HYSYS versions shown in (1) to (12), except that some of the numerical values are slightly different. Especially the coefficients in the m_i expressions (8) and (12) are slightly different.

The parameters in (1) to (12) may be slightly different in different process simulation programs. The alfa (α_i) parameter in equation (7) can be calculated differently. In the PR-Twu and SRK-Twu models the α_i parameter is calculated in a different way (Twu et al., 1991).

Some references to suggest a k_{ij} value for water/CO_2 close to 0.19 are Ziabakshsh-Ganji and Kooi (2012), Li et al. (2014) and Bråthen et al. (2019). Some recommends different k_{ij} values dependent on emphasis on the vapour or liquid phase (Austegard et al., 2006). Some values for water/H_2S from literature are 0.17 and -0.02 optimized for the liquid and vapour phase, respectively (Søreide and Whitson, 1992). A value of 0.105 has been fitted to the binary mixture of water and H_2S (Li et al., 2014; Ziabakshsh-Ganji and Kooi, 2012).

3 Process Description and Simulation Specifications

3.1 Process description of raw biogas compression
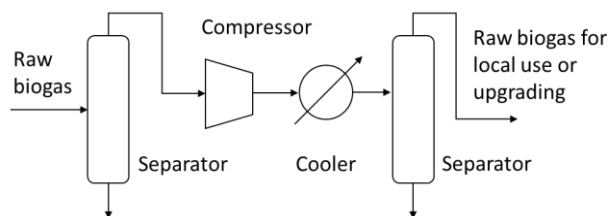
Figure 1. A traditional raw biogas compression process

The principle for a traditional raw biogas compression process is shown in Figure 1.

The actual process is discussed in more detail in Hovland (2017), Øi and Hovland (2018) and Bråthen et al. (2019). Hovland (2017) has suggested to compress the gas to a high pressure, typically above 100 bar, and transport it to a facility for upgrading to biomethane (almost pure methane).

As mentioned in Øi and Hovland (2018) and Bråthen et al. (2019), condensation during compression is regarded to be a problem, and should be avoided.

3.2 Simulation specifications

In earlier work, process simulations have been performed for different conditions relevant for biogas production as in Øi and Hovland (2018) and Bråthen et al. (2019). The models PR, SRK, TST, PR-Twu and SRK-Twu were used. For all the conditions, calculations with the default parameters (especially the k_{ij} for water/CO_2) were used. Simulations with other k_{ij} values have also been performed. Phase envelopes showing the dew and bubble point curve for a temperature and pressure range have been calculated. In the dry gas cases, the HYSYS 2-phase option was selected for phase envelope calculations. In the cases including water and H_2S, the ComThermo 3-phase option was selected.

As a verification, calculations with Aspen Plus and the Stryjek-Vera (1986) model were performed. In Aspen Plus the Peng-Robinson and RKSoave models were selected. The B and D cases are referring to Øi and Hovland (2018) and simulated in Bråthen et al. (2019).

In earlier simulations (Case B) dry biogas with 40 mol-% methane and 60 mol-% CO_2 starts at 37 °C and 1 bar, is cooled to 10 °C and is compressed to 64 bar. In this work, 1 mol-% H_2S is added, and the mol-% methane is reduced to 39.

In earlier simulations (Case D), 59.9 kmol/h methane, 40 kmol/h CO_2 and 0.1 kmol/h water is mixed at 37 ℃ and 1 bar, cooled to 10 °C, and then compressed to 64 bar. In this work, 1 mol-% H_2S is added, and the mol-% methane is reduced to 58.9.

4 Process Simulation, Results and Discussion

4.1 Simulation of dry methane/CO_2 mixture including H_2S (Case B)

An Aspen HYSYS flow-sheet model of the process for the base case simulation is presented in Figure 2.
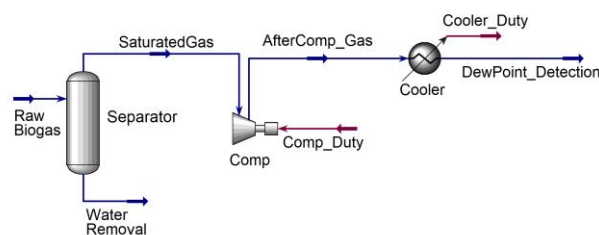
Figure 2. Aspen HYSYS flow-sheet for compression and cooling

In earlier calculations without H_2S, Case B was of interest because a 40 % methane and 60 % CO_2 has a dew point close to 0 °C. Earlier evaluations from Øi and Hovland (2018) and Bråthen et al. (2019) have shown that below 60 mol-% CO_2, no condensation should appear if the temperature is kept above -3 °C.

Table 1. Dew point at 64 bar, cricondenterm and cricondenbar for a mixture of 39 mol-% methane, 60 mol-% CO_2 and 1 mol-% H_2S (Case B)

Model	T_{DEW}	T_{CRIC} (°C)	P_{CRIC} (bar)
PR Hysys	-4.4	-0.5	89.5
SRK Hysys	-4.3	-0.2	88.4
TST Hysys	-2,3	-1.3	82.6
PR-Twu Hysys	-5,3	-1.5	90.0
SRK-Twu Hysys	-4,8	-0,7	90.5
PR Aspen Plus	-4.3	-0.7	88.1
SRK Aspen Plus	-1.8	2.1	82.8

The results in Table 1 are similar to the results from Øi and Hovland (2018) and Bråthen et al. (2019). The results in Aspen HYSYS and Aspen Plus are not identical due to slightly different model parameters. The dew points are typically 1 K and for all the cases less than 3 K higher than the calculations for mixtures without H_2S. These deviations are slightly less than the deviation in calculated dew point between different models used.

Also when H_2S is included, the calculated cricondenterms with different models have a maximum deviation of less than 2 K. It is concluded that the results can be expected to be fairly accurate for all the models evaluated. Even with 1 mol-% H_2S, no condensation will appear above 0 °C in a dry biogas with more than 40 mole-% CH_4. This was also the conclusion without H_2S from Hovland (2017), Øi and Hovland (2018) and Bråthen et al. (2019).

The phase envelope from Aspen HYSYS is shown in Figure 3. The dew point curve is to the right. The point with the highest temperature is the cricondenterm. The point with the highest pressure is the cricondenbar. In the critical point for the mixture, the compositions in both phases are equal.

The phase envelope in Figure 3 is very close to the phase envelope in Bråthen et al. (2019) which was calculated without H_2S. This is the same comparison as the comparison of Table 1 which was also compared with simulations without H_2S. Even in the region close to the critical point and at the maximum pressure, the deviation in temperature seems to be less than 2 K.

These simulations indicate that the calculated dew points and phase envelopes for dry biogas including up to 1 mol-% H_2S are reasonably accurate for all the models tested.
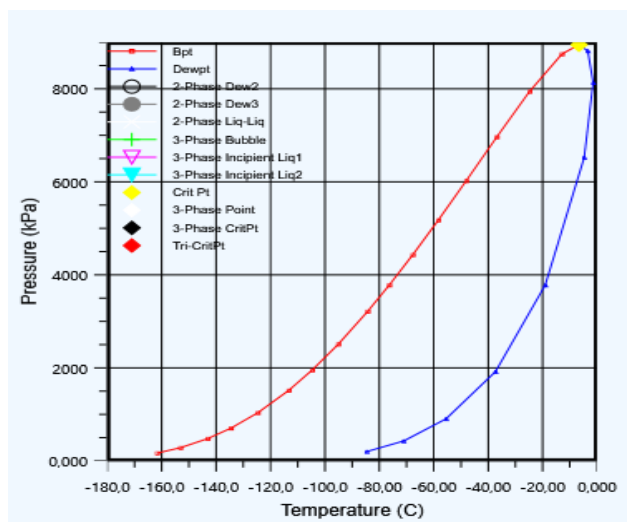
Figure 3. Phase envelope, Peng-Robinson, CH_4=0.39, CO_2=0.6, H_2S=0.01. PR with default k_{ij} values.

4.2 Simulation of compression of a raw biogas including water and H_2S, Case D

The process (Case D) was simulated with water and H_2S included. The simulations are similar to the simulations in Bråthen et al. (2019) and in this work 1 mol-% H_2S is added and the mole fraction of methane is reduced correspondingly. The results are shown in Table 2.

Table 2. Dew point at 64 bar, cricondenterm and pressure at cricondenterm for a mixture of 58.9 mol% CH_4, 40 mol% CO_2, 1 mol% H_2S, 0.1 mol% water, CaseD

Model	T_{DEW} (ºC)	T_{CRIC} (°C)	P_{CRIC} (bar)	X_{H2S}
PR	26.3	-21.1	83.9	0.00062
SRK	26.7	-20.7	83.1	0.00058
PR(k_{ij}=0.105)	26.4	-21.1	83.9	0.000067
PR(k_{ij}=0.105) $k_{H2O/CO2}$=0.19	29.5	-21.1	83.9	0.000063
PRSV	27.0	-22.0	83.5	0.000091
TST+k_{ij}	28.6	-18.1	80.6	0.00062
PRTwu+k_{ij}	28.4	-22.1	83.6	0.00061
SRKTwu+k_{ij}	28.6	-21.5	84.2	0.00058

In the work of Bråthen et al. (2019) other models were also used. The deviation between the models were small. The dew point temperatures in Table 2 were also calculated in Aspen Plus (Bråthen et al., 2019) in simulations without H_2S with other models like PR and SRK. The results were very close to the results using Aspen HYSYS. When using PR and SRK in Aspen Plus the results were identical in the two programs. There were deviations of up to 7 K between the dew point temperature dependent on k_{ij} values.

The calculated cricondenterm values in Table 2 are very much lower than the values in Bråthen et al. (2019). This is probably because the values in this work were calculated without taking water into account. To evaluate the actual cricondenterm values, it is recommended to analyse the phase envelope in **Figure 4** with the similar phase envelope in Bråthen et al. (2019). The dew point curves to the right in the figures are similar, and that is most relevant in this evaluation.

4.3. Fitting of binary parameter for water/H_2S based on literature data.

In (Bråthen et al., 2019) the binary parameter (k_{ij}) for water/CO_2 was varied to give a good fit to experimental dew point data for typical biogas compositions. For mixtures with a methane/CO_2 mole fraction ratio of 50/50 and 70/30, an average value of 0.19 was calculated. This was recommended (Bråthen et al., 2019) to use for calculating dew points in biogas mixtures.

When adding up to 1 mol-% H_2S to a biogas mixture, the results in this work indicate that the dew point temperature is influenced to a low degree by varying the binary parameters (k_{ij} for water/H_2S). The solubility of H_2S in water is however considerably influenced by the k_{ij} for water/H_2S.

In Table 2, the models giving similar H_2S solubility, gave a ratio of the partial pressure of H_2S (= 0.64 bar) and mole fraction H_2S between 1030 and 1100 bar. This can be seen as a Henry's constant, but at the pressure of 64 bar, the gas is not ideal, so as a Henry's constant, it is inaccurate. In Chapoy et al. (2005) similar ratios at 25 °C can be calculated to values between 540 and 550 bar (approximately the H_2S Henry's constant) at total pressure between 5 and 7 bar in a water/H_2S system. The Henrys constant at 25 °C from NIST (2020) is 560 bar (0.10 mol/(kg·bar)).

To compare calculated Henry's constants with literature data, simulations with 1 mol-% H_2S and 99 mol-% water were performed. At 25 ℃, the ratio was calculated to 670 bar with k_{ij} = -0.034 (default) and to 6300 bar with k_{ij} = 0.106. This indicates that the default k_{ij} give a Henry's constant close to literature data at a low pressure. Using a k_{ij} value of 0.106 which is cited in some literature sources (Chapoy et al., 2005; Li et al., 2014) will give a wrong H_2S solubility by an order of magnitude.

The k_{ij} for the water/H_2S binary was adjusted to fit the Henry's constant (547 bar) at 25 ℃ and the result was –0.036. When using the PR and SRK models in Aspen HYSYS and Aspen Plus, the default temperature dependent k_{ij} values give so reasonable results that it is not recommended to change the default k_{ij} values.

4.4. Phase envelope calculations

The phase envelope for PR, k_{ij}=0.19 for water/CO_2 and default k_{ij} for water/H_2S, is shown in **Figure 4**.
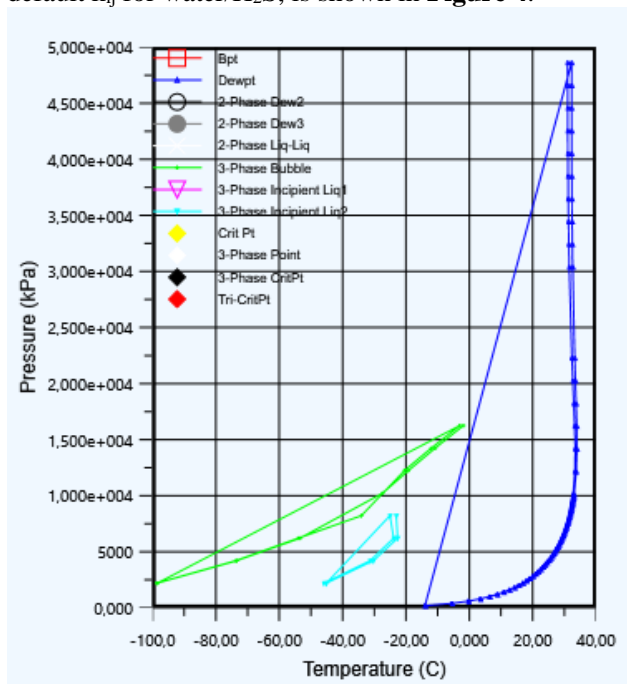
Figure 4. Phase envelope for PR model, 58,9 mol% CH_4, 40 mol% CO_2, 1 mol% H_2S, 0.1 mol% water: k_{ij} =0.19 for water/CO_2, k_{ij} = default for water/H_2S

The phase envelope in **Figure 4** can be compared to a similar phase envelope from Bråthen et al. (2019) without H_2S included. The dew point line to the right in the figure deviate with less than about 2 K for the whole pressure range, compared to the figure without H_2S from 2019. This gives the conclusion that the inclusion of H_2S only slightly influence the dew point temperature calculations. **Figure 4** shows however that there are some additional curves to the left in the figure. These are however at temperatures below 0 ℃. In a raw biogas mixture, temperatures below 0 °C would give the risk of ice formation and hydrate formation. The temperature range below 0 ℃ is not evaluated in this work.

In general, it is expected that the uncertainty in dew point calculations increases when the pressure increases and the mixture is close to the critical point which is about 70 bar (Øi and Hovland, 2018; Bråthen et al., 2019). The range with an uncertainty less than 4 K in calculated dew point with PR or SRK with one constant k_{ij} value (0.19) is for the range of temperatures 0-40 ℃, pressures up to 70 bar and CH_4 concentration above 30 mol-%. This uncertainty is not significantly increased by a H_2S content up to 1 mol-%.

5 Conclusion

Specified mixtures of raw biogas with and without water have been simulated at different temperatures and pressures. The effect of adding up to 1 mol-% H_2S to the mixtures have been evaluated. Dew points have been calculated with different models and with different binary parameters. For some conditions, phase envelopes have been calculated and different models have been compared.

For mixtures of methane and CO_2 with up to 1 mol-% H_2S (a high value for biogas), the different models gave similar results. Under normal ambient temperatures (above 0 °C), a mixture with more than 40 mol-% methane will not give any condensation. When the H_2S increased from 0 to 1 mol-%, the dew point temperature increased with typically 1.0 K. A phase envelope for biogas with 1 mol-% H_2S is only slightly different from an earlier calculated phase envelope for biogas without H_2S.

When raw biogas is cooled or compressed, water will condense first. Some H_2S will dissolve in the water, and the amount of calculated H_2S dissolved in water varied significantly with the model and on the selected binary parameter for water/H_2S. Using the Peng-Robinson model with binary coefficients fitted to H_2S solubility either in the vapor or the liquid phase, gave results for the H_2S solubility in water with a factor of 10 in difference. For biogas simulation, it is recommended to select a binary parameter (fitted to -0.036 at 25 ℃ for PR) that fits experimental data for H_2S solubility in water.

References

A. Aasen, M. Hammer, G. Skaugen, J. P. Jakobsen and Ø. Wilhelmsen. Thermodynamic models to accurately describe the PVTxy-behaviour of water/carbon dioxide mixtures, *Fluid Phase Equilibria*, 442:125-139, 2017.

N. E. Ahmad, M. Mel and N. Sinaga. Design of Liquefaction Process of Biogas using Aspen HYSYS Simulation. *Journal of Advanced Research in Biofuel and Bioenergy*, 2:10-15, 2018.

S. Z. S. Al Ghafri, E. Forte, G. C. Maitland, J.J. Rodriguez-Henriquez and J. P. M. Trusler. Experimental and Modeling Study of the Phase Behaviour of (Methane + CO2 + Water) Mixtures. *Journal of Physical Chemistry*, 118:14462-14478, 2014.

A. Austegard, E. Solbraa, G. de Koeijer and M. J. Mølnvik. Thermodynamic models for calculating mutual solubilities in H_2O-CO_2-CH_4 mixtures. *Trans IChemE, Part A, Chem. Eng. Res. Des.,* 84(A9):781-7946, 2006.

T. Bråthen, L. E. Øi and J. Hovland. Simulation of Dew Points in Raw Biogas Using PR and SRK Equations of State. In *Linköping Electronic Conference Proceedings SIMS 60, pages 31-36, 2018*. doi: 10.3384/ecp20170112

A. Chapoy, A. H. Mohammadi, B. Tohidi, A. Valtz and D. Richon. Experimental Measurements and Phase Behavior Modeling of Hydrogen Sulfide-Water Binary System. *Ind. Eng. Chem. Res.,* 44:7567-7574, 2005.

J. Hovland. Compression of raw biogas – *A feasibility study*. Tel-Tek report 2217020-1, 2017. Available on https://www.biogas2020.se/wp-content/uploads/2017/06/2217020-1compressionrawbiogas.pdf

C. Jarne, S. T. Blanco, M. A. Gallardo, E. Rauzi, S. Otin and I. Valesco. Dew Points of Ternary Methane (or Ethane) + Carbon Dioxide + Water Mixtures: Measurements and Correlation. *Energy & Fuels*, 18:396-404, 2004.

L. N. Legoix, L. Ruffine, J. P. Donval and M. Haeckel. Phase Equilibria of the CH4-CO2 Binary and the CH4-CO2-H2O Ternary Mixtures in the Presence of a CO2-Rich Liquid Phase. *Energies*, 10(2034):1-11, 2017. Doi:10.3390/en10122034.

J. Li, L. Wei and X. Li. Modeling of CO2-CH4-H2S-brine based on cubic EOS and fugacity-activity approach and their comparisons. *Energy Procedia,* 63:3598-3607, 2014.

J. Longhi. Phase equilibria in the system CO2-H2O I: New equilibrium relations at low temperatures. *Geochimica et Cosmochimica Acta*, 69(3): 529-539, 2005.

NIST webbook (Link downloaded 27.6.2020) https://webbook.nist.gov/cgi/cbook.cgi?ID=C7783064&Mask=10#Solubility NIST

D. Peng and D. B. Robinson. A New Two-Constant Equation of State. *Industrial & Engineering Chemistry Fundamentals*, 15(1):59-646, 1976.

L. Pellegrini. Biogas to liquefied biomethane via cryogenic upgrading technologies. *Renewable Energy,* 124:75-83, 2018.

G. Perez, C. Coquelet, P. Paricaud and A. Chapoy. Comparative study of vapour-liquid equilibrium and density modelling of mixtures related to carbon capture and storage with the SRK, PR, PC-SAFT and SAFT-VR Mie equations of state for industrial uses. *Fluid Phase Equilibria*, 77:440, 2017.

R. Privat, F. Mutelet and J. Jaubert. Addition of the Hydrogen Sulfide Group to the PPR78 Model (Predictive 1978, Peng-Robinson Equation of State with Temperature Dependent k_{ij} Calculated through a Group Contribution Method). *Industrial & Engineering Chemistry Research,* 47:10041-10052, 2008.

R. Privat and J. N. Jaubert, *Predicting the Phase Equilibria of Carbon Dioxide Containing Mixtures Involved in CCS Processes Using the PPR78 Model*. InTech, 2014. Available on http://dx.doi.org/10.5772/57058.

G. Soave. Equilibrium constants from a modified Redlich Kwong equation of state. *Chemical Engineering Science*, 27:1197-1203, 1972.

N. Spycher, K. Pruess and J. Ennis-King. CO2-H2O mixtures in the geological sequestration of CO2. I. Assessment and calculation of mutual solubilities from 12 to 100 ºC and up to 600 bar. *Geochimica et Cosmochimica Acta*, 67 (16):3015-3031, 2003.

R. Stryjek and J. H. Vera. PRSV – An Improved Peng-Robinson Equation of State with New Mixing Rules for Strongly Nonideal Mixtures. *The Canadian Journal of Chemical Engineering*, 64:334-340, 1986.

I. Søreide and C. H. Whitson. Peng-Robinson predictions for hydrocarbons, CO2, N2, and H2S with pure water and NaCl brine. *Fluid Phase Equilibria*, 77:217-240, 1992.

Y. Tan, W. Nookuea, H. Li, E. Thorin and J. Yan. Cryogenic technology for biogas upgrading combined with carbon capture – a review of systems and property impacts. *Energy Procedia,* 142:3741-3746, 2017.

C. H. Twu, D. Bluck, J. R. Cunningham and J. E. Coon. A Cubic Equation of State with a New Alpha Function and a New Mixing Rule. *Fluid Phase Equilibria*, 69:33-50, 1991.

Z. Ziabakhsh-Ganji and H. Kooi. An Equation of State for thermodynamic equilibrium of gas mixtures and brines to allow simulation of the effects of impurities in subsurface CO_2 storage. *International Journal of Greenhouse Gas Control*, 2012. doi: 10.1016/j.ijggc.2012.07.025

L. E. Øi and J. Hovland. Simulation of Condensation in Compressed Raw Biogas Using Aspen HYSYS. In *Linköping Electronic Conference Proceedings SIMS 59,* pages 31-36, 2018. doi: 10.3384/ecp1815331

Simple modelling approach using Modelica for microbial electrosynthesis

Gamunu Samarakoon[1]   Dietmar Winkler[2]   Vasan Sivalingam[1]   Carlos Dinamarca[1]   Rune Bakke[1]

[1]Department of Process, Energy and Environmental Technology, [2]Department of Electrical Engineering, Information Technology and Cybernetics, University of South-Eastern Norway, Porsgrunn 3918, Norway,
`{gamunu.arachchige, dietmar.winkler, vasan.sivalingam,`
`carlos.dinamarca, rune.bakke@usn.no}`

Abstract

This study intends to develop a simple mathematical model that contributes to the integration of Microbial Electrosynthesis (MES) in AD to reduce CO_2 to CH_4. Open-source modelling language Modelica was used to build the model. The MES internal resistances are important parameters for the model and an Electrochemical Impedance Spectroscopy (EIS) experiment was employed to estimate the resistances and distinguish the contribution from each resistance element. The model preliminary simulations show that it is possible to determine the voltage required to keep the potential difference across the cathode biofilm within optimal conditions. The system is sensitive to effects of biofilm development on electron transfer at both electrodes, which implies effects on the electrons from anode to cathode (i.e. electric current). The model will be a useful tool for extrapolating experimental results and to enhance our understanding of MES.

Keywords: microbial electrosynthesis, bio-methane, CO_2 reduction, Modelica

1 Introduction

Microbial electrosynthesis is a novel technology for chemical synthesis of desired product through chemical reaction catalysed by microorganisms and powered by electric energy (Rabaey and Rozendal, 2010). MES occurs in microbial electrolysis cell. The working electrode (WE) of the microbial electrolysis cell is in general the cathode at which the reduction half reaction is controlled by the potential to achieve the desired product.

Recently, MES for biogas upgrading (Nelabhotla and Dinamarca, 2018) by CO_2 (carbon dioxide) reduction to CH_4 (methane) has received attention. A typical biogas reactor (anaerobic digester; AD) produces biogas containing 50-70 % CH_4 and 30-50 % CO_2, implying low calorific value (Angelidaki et al., 2018). Biogas upgrading by MES should preferably use electricity from renewable sources and may thereby also serve as a way of storing renewable surplus electricity as CH_4

(Geppert et al., 2016). Other valuable chemical products from CO_2 may also be obtained by adjusting the MES cathode potential (Schievano et al., 2019).

The conversion of CO_2 to CH_4 occurs at the cathode through direct electron transfer (Table 1, eq. (1)) or indirectly via production of intermediates. The conversion of CO_2 to CH_4 with intermediate production of hydrogen (H_2) follows two steps: protons reduction to H_2 (Table 1, eq. (2.1)) and then the produced H_2 is used as electron donor for biological CO_2 reduction to CH_4 (Table 1, eq. (2.2)). CH_4 production is also possible from acetate (CH_3COO^-) produced bio-electrochemically via CO_2 reduction.

Table 1. Chemical reactions at cathode with standard potential (S.P.) [4] vs. Normal Hydrogen Electrode (NHE)

Reactions at cathode	S.P.[V]	
$CO_2 + 8H^+ + 8e^- \rightarrow CH_4 + 2H_2O$	-0.24	(1)
$8H^+ + 8e^- \rightarrow 4H_2$	-0.41	(2.1)
$CO_2 + 4H_2 \rightarrow CH_4 + 2H_2O$		(2.2)

The standard potential of CH_4 production via direct electron transfer (- 0.24V vs. NHE) is lower than that of an indirect electron transfer (- 0.41V vs. NHE) (Table 1), implying that direct electron transfer is the more energy efficient path. The cathode potential also influences current (electron flow) and hence the CH_4 production rate (Geppert et al., 2016) and is the key parameter that determines mechanism of electron transfer.

The protons (H^+) and electrons (e) needed for the reduction reaction at the cathode are generated at the anode, such as by oxidizing water eq. (3) or easily degradable short-chain volatile fatty acid (VFA) such as acetate eq. (4). Oxidation of these components may have some downsides. Thus, other possible oxidation reactions, e.g., ammonium oxidation (Sivalingam et al., 2020), may be relevant.

$$2H_2O \rightarrow 4H^+ + O_2 + 4e^- \qquad 0.82 \text{ V vs. NHE} \quad (3)$$

$$CH_3COO^-H^+ + 2H_2O \rightarrow 2CO_2 + 8H^+ + 8e^-$$

-0.28 V vs. NHE   (4)
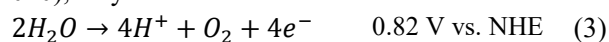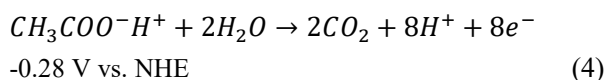
The minimum voltage required to bring about electrolysis can be estimated from the standard reduction potentials of desired reaction and is called thermodynamic cell voltage eq. (5).

$$E^0_{cell} = E^0_{cath} - E^0_{anod} \qquad (5)$$

However, the applied cell voltage consists of not only the thermodynamically calculated cell voltage, but also internal energy losses or over-potentials. These internal losses originate from three sources; activation over-potential (η_{act}, related to the rates of electrode reaction), concentration over-potential (η_{conc}, related to mass transfer limitations of chemical species transported to or from the electrode), ohmic over-potential (η_{ohm}, related to the resistance to the flow of ions in the electrolyte and to the flow of electrons through the electrode material) (Picioreanu et al., 2007). The activation and concentration over-potentials incur separately both at cathode and anode, see eq. (6).

$$V_{cell} = \left(E^0_{cath} - \eta_{C,act} - \eta_{C,con}\right) - \left(E^0_{anod} - \eta_{A,act} - \eta_{A,conc}\right) - \eta_{ohm} \qquad (6)$$

1.1 Electrochemical Impedance Spectroscopy and Equivalent Electrical Circuit

The electrochemical balance eq. (6) expressed as the potential drops at each electrode are influenced by the resistances to the bioelectrochemical reactions and electrochemical processes in MES. These resistances can be converted into an Equivalent Electrical Circuit (EEC) (Jiya et al., 2018). Figure 1 shows a typical EEC of an electrode, where C is the capacitance of the electrical double layer (between the electrode and the electrolyte solution), R_p represents charge transfer resistance originating from MES reactions (related to the activation energy of the electrodes) and R represents resistance by the electrolyte solution. The electrochemical impedance spectroscopy (EIS) has been used to characterize electrochemical system by means of EECs (Wagner, 2002).
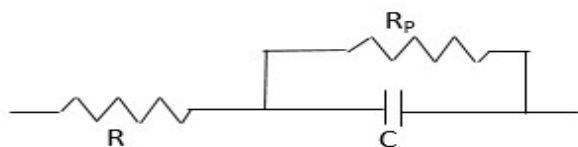
Figure 1: Classical equivalent circuit of a double layer capacitor showing its three basic characteristics: The internal resistance (R), its Capacitance (C), and self-discharge resistance (Rp) (Jiya et al., 2018).

EIS characterizes the response of an electrical circuit to alternative current (AC) or voltage. For a certain amplitude and frequency of applied AC signal, the circuit responds with a particular amplitude of alternating current at the same frequency. This response is quantified as impedance based on Ohm's law (E = I Z, where E and I are AC voltage and current respectively and Z is the impedance). The impedance depends on the frequency of the applied signal (voltage) and the time shift (or phase shift) between the input and output signals.

This technique can be applied to electrochemical cells (in this case MES). An AC signal is applied to the cell and the corresponding response (the impedance) depicts the resistance to charge flow. The charge flow depends on the bio electrochemical reactions and processes in the cell. The different frequencies of the applied AC voltage can distinguish the different processes that have different time scales. The lower frequencies are corresponding to the slow processes such as diffusion and slow electrochemical reactions. The higher frequencies are corresponding to the faster processes. At higher frequencies we can approximate the impedance due to the flow of ions in the electrolyte (i.e., the resistance of the electrolyte solution). Then the impedance response is mainly due to the charge and discharge of the double layer (capacitance).

The impedance data is generated in the form of a spectrum typically on a plot called Nyquist plot. The response impedance Z is composed of a real and imaginary part. The "Nyquist Plot" represents the real part on X-axis and the imaginary part on Y-axis. Bode plot is also another important graph used in EIS analysis where the absolute values of the frequency response (i.e., impedance) and the phase-shift are plotted vs the applied frequency.

The EEC associated with the corresponding spectrum can be constructed with the amplitude of the circuit elements using computer aided tools.

1.2 Aims

This study intends to develop a simple mathematical model that contributes to the integration of MES in AD to reduce CO_2 to CH_4. The experimental work on this is so far limited and will require large efforts to test wide variety of operational conditions, while mathematical modelling can extrapolate on such results and enhance our understanding of MES. In this case, our aim is to develop a simple model to understand the variation in resistances in a MES system as the biofilm grow and in its fully-grown operation. An EIS experiment is employed to estimate the resistances of the electrolyte solution and the charge transfer and distinguish the contribution from each to the MES internal resistance. The Open-source Modelling language Modelica (Modelica Association, 2017) is used to build the simple model. Typical applications of the models as such also include design of power management systems from lab.- to full-scale reactor setups, and the biofilm

resistance values are important to understand the kinetic behaviours of bio-electrochemical reactions.

2 Materials and methods

2.1 MES reactor and operation

The present MES experiments were performed in a 100 mL Borosilicate glass bottle with a 3-ports Teflon screw cap. The reactor was operated in a semi-continuous mode. A carbon nanotube composite (made in-house, area ~1 cm^2) was used as the cathode. A Graphite rod (L: 152 mm × D: 6.15 mm; Alfa Aesar, Thermo Fisher GmbH, Karlsruhe, Germany) was used as the anode. An Ag/AgCl electrode (+0.209 V vs. SHE; 3 M NaCl, QVMF2052, ProSense, BB Oosterhout, The Netherlands) was used as the reference electrode. The electrodes were connected with titanium wire. The voltage between cathode and reference was kept constant at -0.42 V vs SHE by Potentiostat (Interface 1000B, Gamry instruments, Pennsylvania, USA) with the intention of synthesising CH$_4$.

The reactor was operated at 35°C and 10 days HRT with a fully growth biofilm condition. A synthetic feed prepared according to Kenarsari et al. (2020) at a flow rate of 10 mL/day was used. Anaerobic digester sludge from the local municipal wastewater treatment plant (Knarrdalstrand, Porsgrunn, Norway) was used as the inoculum.

2.2 EIS experiment

The current EIS experiment results were generated by running EIS in a cell with a fully-grown biofilm cathode. This experiment work was type of "lesson-learned" and our first attempt was to estimate the impedance associated with the cathode (i.e., WE). We relied on the EIS instrument capacity available in our lab at the time of the study (We intent to improve the capacity and expand our knowledge targeting a pilot scale MES). The EIS experiment was performed by using a potentiostat (Interface 1000B, Gamry instruments, Pennsylvania, USA). The initial frequency was set to 20 kHz, which is the maximum limit of interface 1000B and stepped down to 0.2 Hz through 10 points/decade ratio. The impedance data is generated in the form of a spectrum on Nyquist plot and the Bode plot. The two plots were used to estimate the electrolyte solution resistance. On the Nyquist plot, the intercept of the spectrum with the X-axis (real part of the impedance) in the high-frequency range gives the electrolyte resistance. The impedance in the high frequency range and zero phase-shift was confirmed using Bode plot.

2.3 Model development

The equivalent electrical circuit as given in Figure 1 can be further simplified depending on the application.

Though the biofilm can be represented as an impedance consisting mainly of both resistance and capacitance, in a MES application we will be purely dealing with direct current flows. This means that the capacitance does not have any significant influence on the electrical behaviour of the reactor. Furthermore, we assumed that the resistance of the anode (connected to the counter-electrode, (CE) of the potentiostat) side is negligible compared to the resistance of the solution and biofilm (for this stage of model development). The electrolyte solution itself can be represented by a resistance that also does not change over time since the ionic strength of the solution will stay almost the same.

Figure 2 shows the developed model consisting of an ideal regulated potentiostat voltage source, including sub-models for the development of electrode biofilms and solution properties. The model was build using the open-source Modelling language Modelica (Modelica Association, 2017).
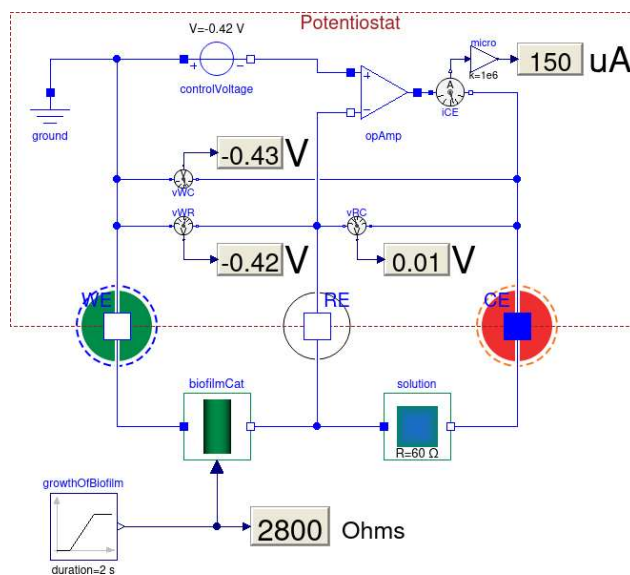
Figure 2: Graphical presentation of potentiostat controlled MES for CO$_2$ reduction to CH$_4$ model implemented in Modelica.

3 Result and discussion

Based on our EIS experiment, we could estimate the electrolyte solution resistance at the cathode side as 60 Ω by observing the Nyquist plot and the Bode plot. Our estimation is based on the frequency response of the highest frequency we could reach with our EIS instrument capacity and the frequency response at zero phase shift. The highest frequency was 20 kHz.
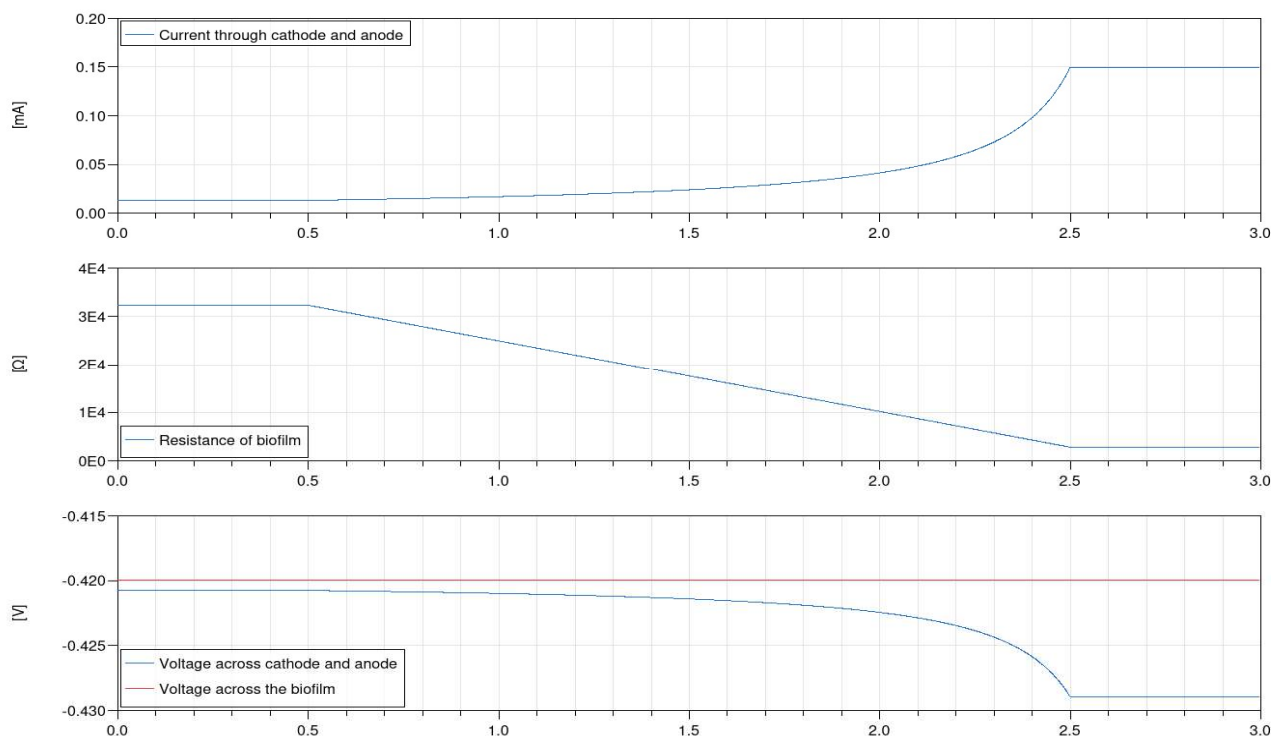
Figure 3: Simulation results for biofilm growth over time in seconds (in reality the growth happened over 21 days).

3.1 Simulation result

The model presented in Figure 2 is used to verify if the voltage level between anode and cathode does not exceed a certain set value (above which undesirable oxidation reactions can occur at the anode or CE) whilst still keeping the voltage across the biofilm at the optimal value of V_biofilm = -0.42 V at the cathode (WE) for CH$_4$ production.

The resistance between anode and solution is neglected for this stage of model development. The resistance of the solution (R_solution) itself has been experimentally determined to be = 60 Ω. Also, from measurements obtained from MES experiments, we could find that at the beginning of the biofilm growth (i.e., abiotic condition), the current flow was 13 μA and after the growth period of 21 days the current went up to 150 μA. Those measurement were taken whilst the potential across the biofilm was always kept at a constant, (i.e. V_biofilm= -0.42 V).

According to Ohm's law this translates to a biofilm resistance of about R_biofilm = 32 kΩ at the beginning of the growth period and 3 kΩ once the biofilm was fully established.

These values were then used in the simulation model in order to determine how high the voltage potential between anode and cathode will get when the voltage drop across the biofilm is kept constant whilst the resistance of the biofilm itself is getting smaller as it is growing.

Figure 3 is showing the simulation results for a change of resistance of the biofilm over time. The

maximum voltage drop in this case does not exceed V_biofilm ≤ 0.43 V. So, the model simulations show that it is possible to determine the voltage required to keep the potential difference across the cathode biofilm within optimal conditions. The system is sensitive to effects of biofilm development on electron transfer at both electrodes, which implies effects on the electric current flow in the cell. This is useful for experimental design.

4 Conclusions and future development

- A functional mathematical model for the integration of MES in AD to reduce CO$_2$ to CH$_4$ is developed in Modelica. It is useful in extrapolating experimental results and to enhance our understanding of MES.

- The cathode resistance was reduced by a factor ten when biofilm was established, compared to the clean electrode. This implies an efficient electron transfer and utilization in the cathode where biofilm develops to reduce CO$_2$ to CH$_4$

- The simulations show that it is possible to determine the voltage required to keep the potential difference across the cathode biofilm within optimal conditions (e.g. V_biofilm ≤ 0.43 V for CO$_2$ to CH$_4$ reduction).

4.1 Further development

We expect to modify this model to better understand impedance variation in the MES system from the abiotic

condition (without biofilm), to the fully-grown biofilm operation (i.e., steady state). Hence, the model can predict the cell voltage (across the WE and CE), whilst keeping the optimum potential at the biocathode (or a WE) at different stage of MES operation. EIS experiments will also be carried out accordingly.

References

Irini Angelidaki, Laura Treu, Panagiotis Tsapekos, Gang Luo, Stefano Campanaro, Henrik Wenzel, and Panagiotis G. Kougias. Biogas upgrading and utilization: Current status and perspectives. *Biotechnology Advances, 36*(2): 452-466, 2018. doi:https://doi.org/10.1016/j.biotechadv.2018.01.011.

Florian Geppert, Dandan Liu, Mieke van Eerten-Jansen, Eckhard Weidner, Cees Buisman, and Annemiek ter Heijne. Bioelectrochemical Power-to-Gas: State of the Art and Future Perspectives. *Trends in Biotechnology, 34*(11): 879-894, 2016. doi:https://doi.org/10.1016/j.tibtech.2016.08.010.

Immanuel N. Jiya, Nicoloy Gurusinghe, and Rupert Gouws. Electrical Circuit Modelling of Double Layer Capacitors for Power Electronics and Energy Storage Applications: A Review. *Electronics, 7*(11): 268, 2018.

Zahra Nikbakht Kenarsari, Nirmal Ghimire, Rune Bakke, and Wenche Hennie Bergland. (2020). *Thermophilic Anaerobic Digestion Modeling of Lignocellulosic Hot Water Extract using ADM1.* Paper presented at the Proceedings of The 60th SIMS Conference on Simulation and Modelling SIMS 2019, August 12-16, Västerås, Sweden.

Modelica_Association. Modelica®–A Unified Object-Oriented Language for Systems Modeling, Language Specification, Version 3.4, April 2017. *Online: https://www.modelica.org/documents*, 2017.

A. B. T. Nelabhotla and C. Dinamarca. Electrochemically mediated CO_2 reduction for bio-methane production: a review. *Reviews in Environmental Science and Bio/Technology, 17*(3): 531-551, 2018. doi:10.1007/s11157-018-9470-5.

Anirudh Bhanu Teja Nelabhotla and Carlos Dinamarca. Bioelectrochemical CO_2 Reduction to Methane: MES Integration in Biogas Production Processes. *Applied Sciences, 9*(6): 1056, 2019.

Cristian Picioreanu, Ian M. Head, Krishna P. Katuri, Mark C. M. van Loosdrecht, and Keith Scott. A computational model for biofilm-based microbial fuel cells. *Water Research, 41*(13): 2921-2940, 2007. doi:https://doi.org/10.1016/j.watres.2007.04.009.

Korneel Rabaey and René A. Rozendal. Microbial electrosynthesis — revisiting the electrical route for microbial production. *Nature Reviews Microbiology, 8*: 706, 2010. doi:10.1038/nrmicro2422.

Andrea Schievano, Deepak Pant, and Sebastia Puig. Editorial: Microbial Synthesis, Gas-Fermentation and Bioelectroconversion of CO_2 and Other Gaseous Streams. *Frontiers in Energy Research, 7*(110), 2019. doi:10.3389/fenrg.2019.00110.

Vasan Sivalingam, Carlos Dinamarca, Gamunu Samarakoon, Dietmar Winkler, and Rune Bakke. Ammonium as a Carbon-Free Electron and Proton Source in Microbial Electrosynthesis Processes. *Sustainability, 12*(8): 3081, 2020.

N. Wagner. Characterization of membrane electrode assemblies in polymer electrolyte fuel cells using a.c. impedance spectroscopy. *Journal of Applied Electrochemistry, 32*(8): 859-863, 2002. doi:https://doi.org/10.1023/A:1020551609230.

Computational study of CO$_2$ injection at Johan Sverdrup for enhanced oil recovery and storage

Ole Christopher Hansen   Britt M. E. Moldestad

Department of Process, Energy and Environmental Technology, University of South-Eastern Norway, Norway,
Olechr95@gmail.com Britt.Moldestad@usn.no

Abstract

Injection of supercritical carbon dioxide (CO$_2$) for enhanced oil recovery (EOR), plays a vital role to minimize the impact of CO$_2$ emissions. CO$_2$-EOR refers to the oil recovery technique where supercritical CO$_2$ is injected in the reservoir to stimulate oil production from depleted oil fields. CO$_2$-EOR can be used in combination with CO$_2$ storage to mitigate the emissions levels to the atmosphere. The objective of this paper is to perform a computational study of CO$_2$-EOR and storage at the Johan Sverdrup field. The study includes simulations of oil production using the commercial software Rocx in combination with OLGA. Production with inflow control devices (ICD) and autonomous inflow control valves (AICV) shows that AICVs have an oil-to-water ratio of 0.92 compared to 0.39 for ICDs. CO$_2$-EOR in combination with well completion with AICVs, shows improved oil recovery, low water production, and low CO$_2$ reproduction. The simulations and calculations performed in this paper indicate that the Johan Sverdrup field is highly capable for CO$_2$-EOR and storage.

Keywords: increased oil recovery, Johan Sverdrup field, CO$_2$ EOR, inflow control devices, OLGA/Rocx simulation

1 Introduction

The oil industry uses several different technologies to increase the oil recovery from new and existing reservoirs. Enhanced Oil Recovery (EOR) can be used in new, existing, and also closed wells to increase the oil production. A possible method for EOR is injection of CO$_2$ as a supercritical fluid. Supercritical CO$_2$ has a significantly higher density than CO$_2$ gas, whereas the viscosity is about the same for the two phases. The positive side effect of injecting CO$_2$ in reservoirs is the storage possibilities.

Figure 1 shows the expected remaining resources in some of the most significant fields on the Norwegian continental shelf (NCS). According to the Resource report published by the Norwegian Petroleum Directorate in 2019, the number of recoverable resources as of 31st December 2018, was estimated to be 15.6 billion standard cubic meters (Sm^3) of oil

equivalent (o.e.), including all the resources previously produced (Norwegian Petroleum Directorate, 2019). The Johan Sverdrup field started up during the autumn 2019 and is considered as the most prominent field development for the past 20 years.
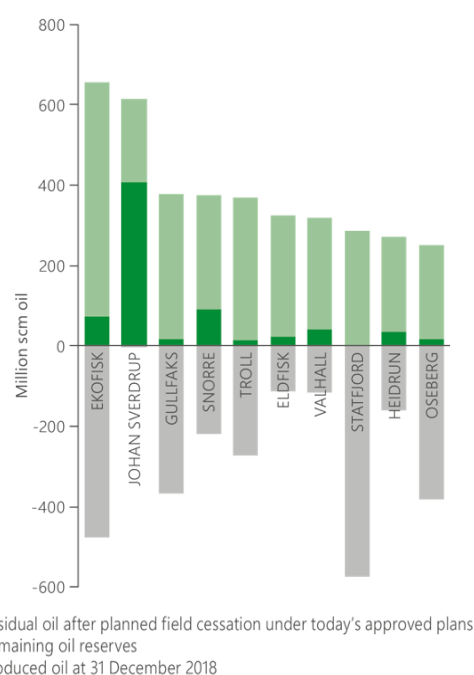
Figure 1. Distribution of oil resources and oil reserves in the 10 largest oil fields on the Norwegian shelf.

The improvements developed concerning horizontal wells in the last decades have made it possible to use long and multilateral wells. Long horizontal and multilateral wells are used to obtain maximum reservoir contact, which results in better production efficiency. Horizontal wells have a frictional pressure drop from the toe to the heel. The heel is located where the well bends from vertical to horizontal, and the toe is the endpoint of the horizontal well. As can be seen in Figure 2, the pressure difference between the reservoir and the well is significantly higher in the heel (ΔP_h) compared to the

(ΔP_t). This is called the heel-to-toe effect. The pressure difference between reservoir and well is called drawdown, and indicates the driving force for production.
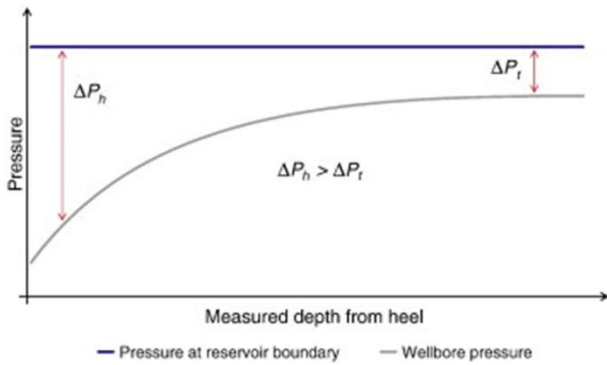
Figure 2. Frictional pressure drop and pressure difference (Birchenko et al., 2010)

The motivation for combining CO_2-EOR and storage could open possibilities for long term storage capacity in the future. Increased yield by CO_2-EOR and storage could appear as appealing to customers from an economic standpoint. Storing CO_2 would help towards Norway's goals of becoming a low carbon society by 2050.

2 Inflow control

This section includes the description of two different types of inflow control devices.

2.1 Inflow control devices

Different types of inflow control devices are installed as a part of the well completion to help optimizing the production by adjusting the flow rate along the wellbore. The purpose of these devices is to delay the breakthrough of either gas or water. Delaying the breakthrough is typically done by reducing the annular velocity across each section, such as the heel of a horizontal well. Multiple ICDs are installed with different diameters in each zone in the well. The diameter is smallest near the heel since the drawdown in this area is significantly higher than in the toe section. By installing ICDs with small diameter in the heel section, the flow is reduced and becomes equal to the flow in the toe section. Thus, an even inflow is achieved from all parts of the well, and early breakthrough of water or gas in the heel is avoided. Figure 3 shows how the implementation of ICDs will even out the breakthrough of gas and water, thus increasing production from each well.

The principle of the nozzle ICD is based on the following equations (Aakre, 2017):

$$\Delta P = \frac{\rho v^2}{2C^2} = \frac{\rho \dot{Q}^2}{2A_{Valve}^2 C^2} = \frac{8\rho \dot{Q}^2}{\pi^2 D_{Valve}^4 C^2} \qquad (1)$$

$$C = \frac{C_D}{\sqrt{(1-\beta^4)}} = \frac{1}{\sqrt{K}} \qquad (2)$$

$$\beta = \frac{D_2}{D_1} \qquad (3)$$

ΔP is the pressure drop through the nozzle ICD, ρ is the average fluid density, v is fluid velocity through

the nozzle, $\dot{Q}$ is the fluid flow rate through the nozzle, A is the cross-sectional area of the nozzle, C is the flow coefficient, C_D is the discharge rate coefficient, K is the pressure drop coefficient, and D is the diameter of the nozzle. These equations show that the nozzle ICD is independent of the fluid viscosity.
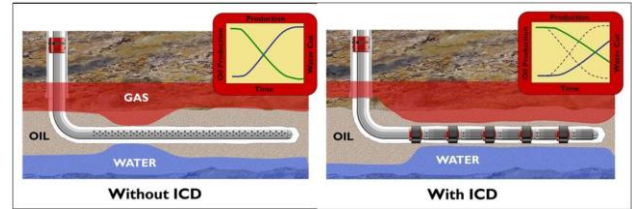
Figure 3. Illustration of water and gas breakthrough without and with ICDs (Halliburton, 2010).

2.2 Autonomous inflow control valves

The autonomous inflow control valve (AICV) can distinguish between fluids based on fluid viscosity and density. The principal operating feature of the AICV is to open for high viscosity fluids and close for low viscosity fluids. This mechanism is controlled by a minor pilot flow that flows parallel to the main flow, and this is shown in Figure 4. The mechanism is described in detail in (Aakre, 2017; Aakre et al., 2013; Kais et al., 2016; Aakre et al., 2018).
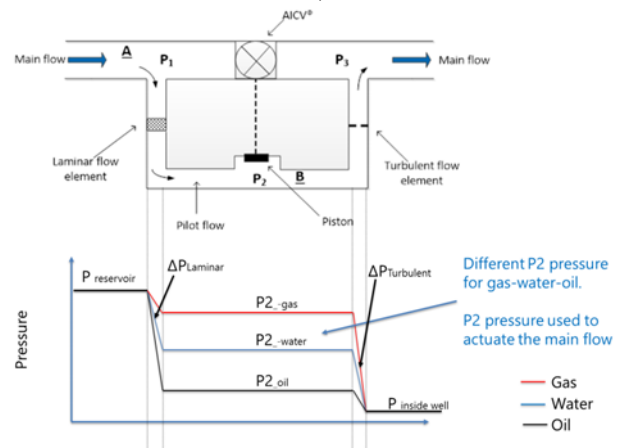
Figure 4. The principle of the AICV (Aakre, 2017).

When the difference in pressure (P_1-P_2) is high, as in the case of oil, the valve stays in the open position. If the pressure difference is low, as in the case of water or gas, the valve will return to the closed position. A closed valve allows for zoning out different sections along the wellbore, thus increasing the recovery of the desired fluid. During production, the AICV is designed to have approximately 99% of the flow rate in the main flow, and when the valve is closed, the minor pilot flow represents the total flow rate through the valve (Aakre, 2017).

The pressure difference can be expressed based on the laminar and turbulent flow elements:

$$\Delta P_{laminar} = f \cdot \frac{L \rho v^2}{2D} = \frac{64}{Re} \cdot \frac{L \rho v^2}{2D} = \frac{32 \mu v L}{D^2} \qquad (4)$$

where f is the laminar friction factor ($64/Re$), ρ is the fluid density, v is the fluid velocity, μ is the fluid viscosity, D is the diameter of the laminar flow element, and L is the length of the laminar flow element.

$$\Delta P_{turbulent} = k \cdot \frac{\rho v^2}{2} \qquad (5)$$

where k is a geometric constant.

3 Material and methods

3.1 Simulation tools

To predict the production of oil and water from a reservoir, simulations have been done in OLGA-Rocx. The OLGA-Rocx module is a dynamic reservoir model designed to model flow rates and pressures near the wellbore. OLGA simulates flow rates and pressures based on the reservoir model defined in Rocx. The simulations are performed over a specified period, and changes are registered over time.

In this study, the conditions from the Johan Sverdrup field in the North Sea were used as a basis for the simulations. As shown in Figure 5, the Johan Sverdrup field is located approximately 150 km west of Stavanger, Norway. Most of the data collected for these simulations are made public by Equinor. However, some values had to be approximated, such as the ratio between horizontal and vertical permeability and the relative permeability.

Figure 5. Geographical location of the Johan Sverdrup field (Equinor, ASA, 2020)

3.1.1 Rocx

The geometry of the reservoir was modeled in Rocx, and is presented in Figure 6. The reservoir is a homogenous sandstone reservoir. The geometry for the simulated reservoir is 1312.5 m in length, 100 m in width and, 40 m in height. 21 grid blocks are defined in the x-direction,

23 grid blocks in the y-direction, and 10 in the z-direction. The well is located 20 m from the bottom, which is indicated by the black circle. The simulated reservoir represents the production area for one well in a large reservoir.
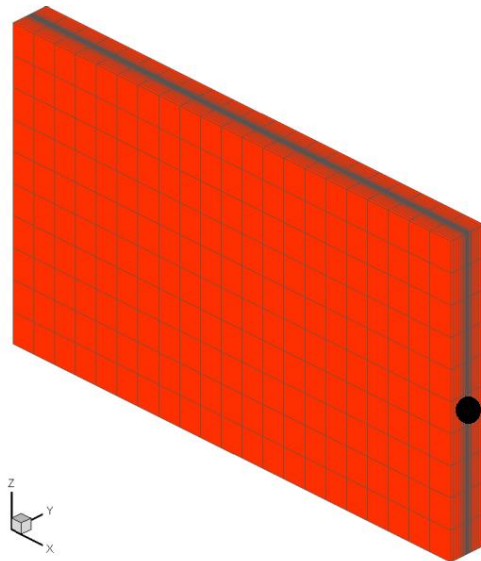
Figure 6. Grid resolution in a reservoir section designed in Rocx. Black dot is the well.

Since the reservoir is homogenous, the porosity and permeability are constant at 0.28 and 5000 mD respectively in the entire reservoir. In these simulations, a ratio between vertical and horizontal permeability of 0.1 is used. The temperature is maintained constant at 83 °C, the water drive pressure is constant at 195 bar, and the wellbore pressure is set to 192 bar in the heel section. The reservoir and fluid properties are presented in Table 1.

Table 1. Reservoir and fluid properties.

Properties	
Oil viscosity	3 cP
Oil density	810 kg/m^3
Reservoir pressure	195 bar
Reservoir temperature	83 °C
Oil specific gravity	0.81
Wellbore pressure (heel)	192 bar
Permeability (x-y-z direction)	5000-5000-500 mD
Porosity	0.28

3.1.2 OLGA

The OLGA simulator is governed by conservation of mass equations for gas, liquid and liquid droplets, conservation of momentum equations for the liquid phase and the liquid droplets at the walls, and conservation of energy mixture equation with phases having the same temperature (Schlumberger Software, 2020).

The OLGA module uses the model developed in Rocx by importing data through the near-well source. To achieve an accurate representation of a wellbore,

with valves, packers, annulus, and production pipe, OLGA requires a "Flowpath" and "Pipeline" (Sund et al., 2017) as shown in Figure 7. For an accurate representation of the wellbore, two separate production pipes need to be defined to account for the annulus and production pipe. "Flowpath" is defined as the production pipe, and "Pipeline" is defined as the annulus. The entire well is divided into 42 equal sections. As shown in the figure, the connection to Rocx occurs through the sources. These sources indicate inflow from the reservoir to the annulus. The inflow control (ICD or AICV), together with the leaks, indicate the flow from the annulus into the pipe. The packers are simulated as closed valves, and their purpose is to isolate different production zones in the well. The packers divide the production pipe into 21 zones.
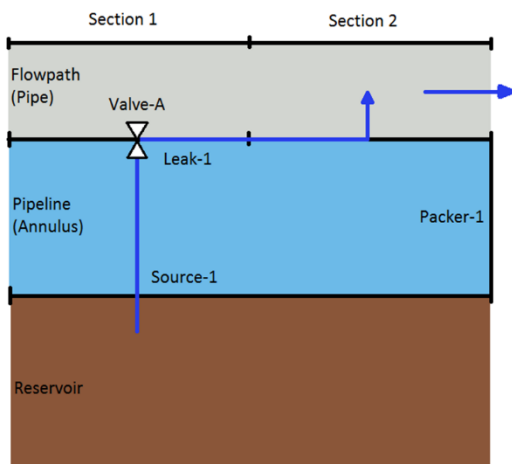
Figure 7. Excerpt of near-well simulation in OLGA.

3.1.3 Simulation set-up for ICD and AICV

Two cases, Case 1 and Case 2, were modelled and simulated. The ICD case (Case 1) was modeled with fully open valves, and in the AICV case (Case 2) the valves were modelled to be adjusted between open and closed based on the nearby fluid properties. In Figure the valve is called VALVE-2 and is either an ICD or an AICV.
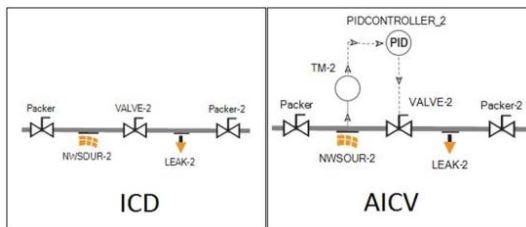
Figure 8. Set-up for one pipe section in OLGA.

There are no options to choose an autonomous inflow control valve in OLGA. The function of an AICV is therefore modeled using a transmitter and a PID regulator. The PID regulator was given a set-point of water cut (WC), and the valve opening was based on the actual WC. Figure 9 shows the outlet of the production pipe (Flowpath) with a choke and a PID regulator, which

regulates with regards to the total flow rate. This system is implemented to account for the maximum capacity of an oil refinery and has a set point at 100 m^3/h.
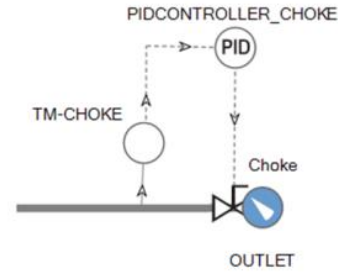
Figure 9. Sketch of the well outlet including a choke with a PID regulator.

3.1.4 Oil production with CO_2 EOR

Case 3, enhanced oil recovery using CO_2 in combination with AICV, is prepared for simulations. Case 3 was simulated with CO_2 already injected in the water phase. To avoid CO_2 short-circuiting between injectors and producers, AICVs were installed along the pipeline. AICVs close off the zones with early breakthrough of CO_2 and water, and ensure that CO_2 will be evenly distributed in the reservoir. The initial oil saturation in the reservoir for Case 1 and 2 is 100%, whereas the oil saturation for Case 3 is 50%.

The relative permeability changes when CO_2 is injected to a reservoir, and the oil is getting more mobile due to oil swelling, reduced viscosity and decrease of the interfacial tension. The relative permeability curves are adjusted and implemented in Rocx. Figure 10 shows the relative permeability curves for oil and water with and without CO_2.
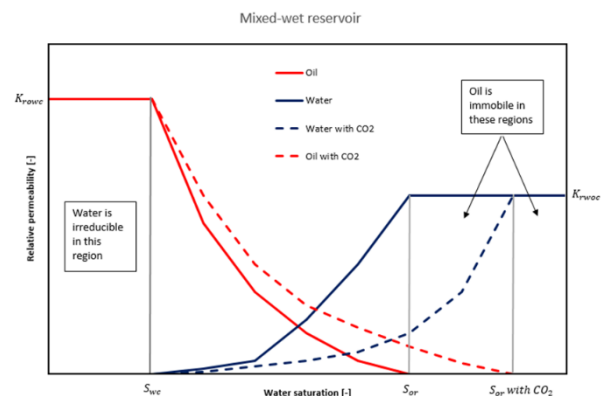
Figure 10. Relative permeability curves for oil and water; with and without CO_2 injection.

The residual oil saturation is reduced from 0.3 to 0.1 when CO_2 is injected to the reservoir. The irreversible water saturation is set to 0.2 in all the cases.

3.2 CO_2 storage

Before injecting supercritical CO_2 ($scCO_2$) into a reservoir or an aquifer, the storage capacity has to be evaluated. The storage capacity can be calculated based

on the model developed by Szulczewski and Juanes (Szulczewski and Juanes, 2009):

$$C_s = \left[\frac{2M\Gamma^2(1-S_{wc})}{\Gamma^2+(2-\Gamma)(1-M+M\Gamma)} \right] \rho_{scCO_2} \phi H W L_{total} \quad (6)$$

where C_s is the storage capacity, M is the mobility ratio, Γ is the trapping coefficient, S_{wc} is the connate water saturation, ρ_{scCO_2} is the density for scCO2, ϕ is the porosity, H is the thickness of the sandstone, W is the length of the injection array, and L_{total} is the total length of the simulated reservoir. The bracketed term in Equation (6) is the storage efficiency and relates the total pore volume to the volume of trapped CO2. The mobility ratio is defined as:

$$M = \frac{1/\mu_w}{K_{rCO_2}^*/\mu_{CO_2}} \quad (7)$$

where μ_w is the viscosity of brine, μ_{CO_2} is the viscosity of scCO2, and $K_{rCO_2}^*$ is the endpoint relative permeability to scCO2. The viscosities used for brine and scCO2 are 0.5 and 0.043 cP, respectively. The trapping coefficient is defined as:

$$\Gamma = \frac{S_{rCO_2}}{1-S_{wc}} \quad (8)$$

where S_{rCO_2} is the residual saturation of CO2. The Szulczewski and Juanes model also includes an equation for the CO2 footprint, which is used to calculate how far the CO2 plume migrates away from the injection array when it is completely trapped.

$$L_{max} = \left[\frac{(2-\Gamma)(1-M(1-\Gamma))}{(2-\Gamma)(1-M(1-\Gamma))+\Gamma^2} \right] L_{total} \quad (9)$$

Figure 11 illustrates the CO2 footprint.
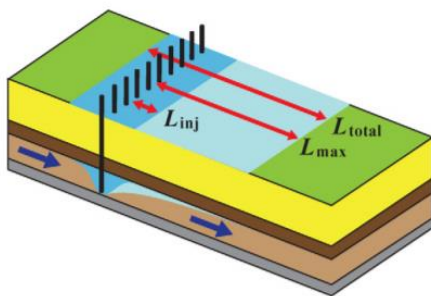
Figure 11. Injection and trapped CO2 footprint (Szulczewski and Juanes, 2009)

The dark blue footprint is injection footprint and is measured as L_{inj}, the light blue footprint is the trapped footprint and is measured as L_{max}. The total plume after it is trapped is L_{total}.

As this is a simple model, it includes several assumptions. The reservoir is assumed to be horizontal, homogenous and isotropic, and the injected plume

migrates with the aquifer flow. The fluid properties, such as viscosity and density, are assumed to be constant. (Szulczewski and Juanes, 2009)

4 Results

4.1 Oil production with ICD and AICV

A simulation model is developed based on the available information from the Johan Sverdrup field in the North Sea. The oil is classified as light oil, because of its high mobility with rather low viscosity and density. Two initial cases were simulated, one with ICDs and one with AICVs. The reservoir has an underlying aquifer, and the AICVs are used to close for water from this aquifer. Figure 12 presents the accumulated water production from the two cases. After about 200 days of production, the set point for the water cut is reached, and the AICVs start to close, to reduce the water production. The ICDs are passive devices, and do not have the ability to close for water. After 600 days, the accumulated water is 2.6 times higher for Case 1 compared to Case 2.
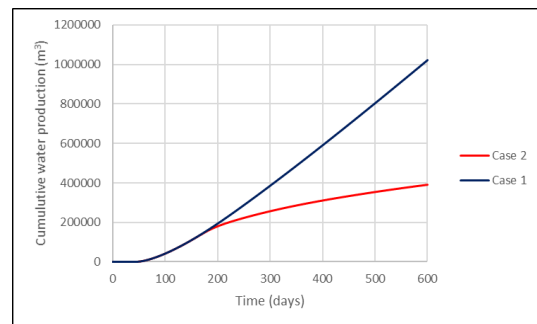
Figure 12. Water production versus time for Case 1 and Case 2.

The oil production from the two cases is compared in Figure 13. During the first 200 days, the oil and water production is equal for Case 1 and Case 2. This is because the AICVs in fully open position have the same inflow area as the ICDs. However, after the initial 200 days, when the AICVs have started to close, the oil production in Case 1 is higher than in Case 2. After 600 days of production, the accumulated oil for Case 1 is about 10% higher than for Case 2.
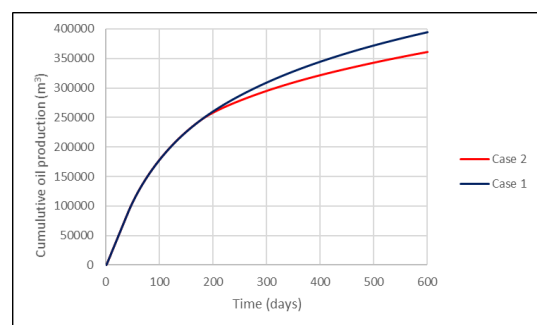
Figure 13. Oil production versus time for Case 1 and Case 2.

4.2 Enhanced oil recovery by injecting CO₂

Case 3, with CO_2 injection, can be considered as a continuation of Case 2. The initial oil saturation is determined to be 50%, which is the average oil saturation in Case 2 after 600 days. The simulation is performed assuming that CO_2 is already injected and well distributed in the reservoir. As the simulation with CO_2 was initiated with higher water saturation and lower oil saturation, it is expected that the flow control valves would begin throttling down sooner than the case with higher initial oil saturations. From Figure 14, it is can be seen how the AICVs close with time. The closure of the AICVs is determined by the water cut, and after about 300-400 days, it seems like all the AICVs are closed. In closed position, the AICVs, in this case, are producing from about 5% of the fully open valve area. After 600 days, the oil production can still be maintained as long as it provides a financial gain. After 600 days, the accumulated oil and water production is about 92000 m³ and 189000m³ respectively, and the average water cut is about 67%.
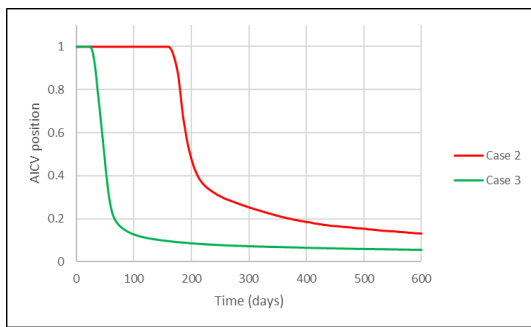
Figure 14. AICV position versus time.

4.3 Comparison of the results

Petroleum companies want a high oil/water ratio for their production chain. Water produced together with the oil results in problems for the oil industry, both economically and environmentally. The water cannot be pumped directly back to the sea, but requires cleaning to comply with national and international environmental policies. Thus, an effort to reduce water production but keep the oil production as high as possible is crucial. Autonomous inflow control devices, such as AICV, are developed to meet this challenge. Figure 15 shows a typical oil production case with ICD (Case 1), and an improved oil production case with the implementation of AICV (Case 2). Wells with ICD completion delay the water breakthrough in the same way as AICV completed wells, but as soon as the breakthrough occurs, the ICDs are not capable of choking the flow, and significant amounts of water are produced together with the oil. After water breakthrough, the AICVs are closed one by one. In a homogeneous reservoir, the drawdown, and thereby the production rate is highest in the heel section, and the earliest water breakthrough will occur in this

section. When zones with water breakthrough are closed off, oil can still be produced from the other zones. Therefore, the AICV case are producing almost the same quantity of oil as the ICD case, but significantly less water.
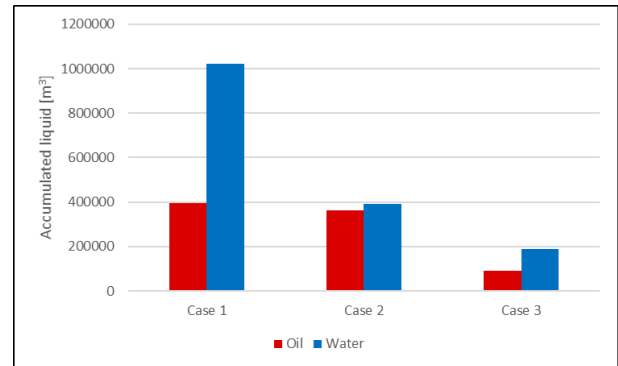
Figure 15. Comparison of accumulated oil and water.

The figure also shows the water and oil produced from the CO_2 injection case (Case 3). CO_2 EOR is mainly used in mature oil fields to increase the oil recovery. Case 3 is therefore run with an oil saturation of 50%. CO_2-injection changes the oil properties, and makes the oil more mobile. It is therefore possible to increase the oil recovery. When CO_2 injection is combined with AICV completion, the oil recovery can be increased without producing large amounts of water. CO_2-EOR is used to extract more oil from the reservoir when normal production is finished. The oil/water ratios for Case 1, Case 2 and Case 3 are 0.39, 0.92 and 0.49 respectively.

4.4 CO₂ storage capacity, Johan Sverdrup

The storage capacity model is used to give an intuitive picture of how much CO_2 can be stored in the Johan Sverdrup field. The area of the Johan Sverdrup field is about 200 km, and the reservoir thickness is set to 40 m. This gives a total volume of $8 \cdot 10^9$ m³. The storage capacity and the CO_2 plume together with the properties of supercritical CO_2 are presented in Table 2. The storage capacity for CO_2 in the Johan Sverdrup field is found to be 49 megaton.

Table 2. CO₂ storage capacity and plume dimensions.

Parameter	Value
ϕ	0.28
S_{wc}	0.3
S_{rCO2}	0.4
k^*_{rCO2}	0.6
μ_{scCO_2}	0.000043 Pa·s
ρ_{scCO_2}	562 kg/m³
Results	
C_s	$4.9 \cdot 10^7$ ton
L_{max}	≈20 km
$L_{injection}$	≈5 km
Efficiency factor	3.9%

Figure 16 presents the distribution of CO_2 in the reservoir and underlying aquifer. The red dot visualize the injection points, the light blue footprints is the trapped footprint, and the dark blue footprints is the injection footprint. It is assumed that the injected CO_2 migrates north of the injection point.
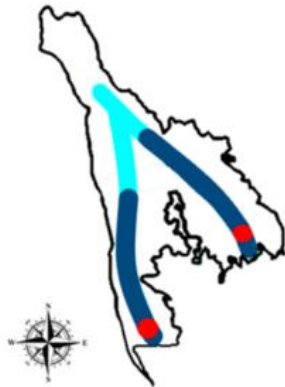
Figure 16. Hydrogeological footprint of CO_2 at Johan Sverdrup.

The information on the aquifer in the Johan Sverdrup reservoir is scarce. The visualized migration of the CO_2 plume from the injection point is therefore a hypothetical scenario, where the CO_2 follows the aquifer flow. The assumption made is that the plume migrates to the north of the injection point.

5 Conclusion

The objective of this paper is to study the CO_2 injection at the Johan Sverdrup field. The study included near-well simulations of oil production and CO_2 injection using the simulation tool OLGA in combination with Rocx. Three different cases were simulated for an intermediate-wet reservoir with high permeability, low oil viscosity, and low oil density. Production with inflow control devices (ICD) and autonomous inflow control valves (AICV) shows that AICVs have an average oil-to-water ratio of 0.92 compared to 0.39 for ICDs. CO_2-EOR in combination with AICVs, shows improved oil recovery, low water production, and low CO_2 reproduction. CO_2 injection into the Johan Sverdrup field increases oil recovery by approximately 33%. Oil produced from Case 1 (ICD), Case 2 (AICV), and Case 3 (CO_2 and AICV) is 395 000 m^3, 361 000 m^3, and 92 000 $m^{3,}$ respectively. CO_2 EOR and storage are heavily dependent on the petrophysical properties of the reservoir. The CO_2 storage capacity for the Johan Sverdrup field is calculated to be 49 megatons, with an efficiency factor of 3.9%. The storage capacity is considered as very promising for the Johan Sverdrup reservoir.

References

H. Aakre. *The Impact of Autonomous Inflow Control Valve on Increased Oil Production and Recovery*, Doctoral thesis, University of South-Eastern Norway (USN), Porsgrunn, 2017.

H. Aakre, B. Halvorsen, B. Werswick, and V. Mathiesen. Smart well with autonomous inflow control valve technology, *SPE 164348-MS*, *SPE Middel East Oil and Gas Show and Exhibition*, Barhain, March 2013.

H. Aakre, V. Mathisen, and B. Moldestad. Performance of CO2 flooding in a heterogeneous oil reservoir using autonomous inflow control. *Journal of Petroleum Science and Engineering*, 167: 654-663, 2018.

V. M. Birchenko, K. M. Muradov, and D. R. Davies. Reduction of the horizontal well's heel–toe effect with inflow control devices, *Journal of Petroleum Science and Engineering*: 244-250, 2010.

Equinor ASA. *Facts about the Johan Sverdrup field*, https://www.equinor.com/en/what-we-do/johan-sverdrup/johan-sverdrup-facts.html

Halliburton. *Inflow Control Devices: Extending the Life of Mature Field Wells*, 30th May 2012. https://halliburtonblog.com/inflow-control-devices-extending-the-life-of-mature-field-wells/

R. Kais, V. Mathiesen, H. Aakre, G. Woiceshyn, A. Elarabi, and R. Hernandez. *First Autonomous Inflow Control Valve (AICV) Well Completion Deployed in a Field under an EOR (Water & CO2 Injection) Scheme*, SPE-181552-MS, 2016.

Norwegian Petroleum Directorate. *Resource report* 2019, https://www.npd.no/globalassets/1-npd/publikasjoner/ressursrapport-2019/resource-report-2019.pdf.

Schlumberger Software ltd., OLGA/ROCX, https://www.software.slb.com/products/olga/olga-wells-management/rocx [Accessed 11th March 2020]

S. M. Sund, T. B. Iversen, M. S. Hansen, J. Bohlin, O. K. Vøllestad, N. C. I. Furuvik, B. M. Halvorsen, Simulation of Oil Production from Homogenous North Sea Reservoirs with Inflow Control using OLGA/Rocx, *Linköping Electronic Conference Proceedings*, 138: 188-195, 2017.

M. Szulczewski and R. Juanes. A simple but rigorous model for calculating CO2 storage capacity, *Energy Procedia,* 1, 2009.

Simulation and Cost Optimization of different Heat Exchangers for CO₂ Capture

Solomon Aforkoghene Aromada[1], Nils Henrik Eldrup[1,2], Fredrik Normann[3], Lars Erik Øi[1]

[1]Department of Process, Energy and Environmental Technology, University of South-Eastern Norway
[2]SINTEF Tel-Tek, SINTEF Industri, Kjølnes Ring 30, 3918, Porsgrunn, Norway
[3]Chalmers University of Technology, 412 96, Göteborg, Sweden
Corresponding author's email address: solomon.a.aromada@usn.no; saromada@gmail.com

Abstract

The industrial deployment of amine-based CO₂ capture technology requires large investments as well as extensive energy supply for desorption. Therefore, the need for efficient cost and economic analysis aimed at CO₂ capture investment and operating costs is imperative. Aspen HYSYS simulations of an 85% CO₂ absorption and desorption process for flue gas from cement industry, followed by cost estimation have been performed. This is to study the cost implications of different plants options. Each plant option has a different lean/rich heat exchanger type. Cost optimisation of the different heat exchangers is also done in this work. Three different shell and tube and two plate and frame heat exchangers have been examined. The minimum CO₂ capture cost of €57.9/ton CO₂ is obtained for a capture plant option having a gasketed-plate heat exchanger with ΔT_{min} of 5 ℃ as the lean/rich heat exchanger. The use of plate and frame heat exchangers will result in considerable CO₂ capture cost reduction.

Key words: simulation, CO2, CCS, heat exchanger, shell and tube, Aspen HYSYS, plate heat exchanger

1 Introduction

There has been increased public concern for mitigation of global warming, which is largely caused by emissions of carbon dioxide (CO₂). Carbon capture and storage (CCS) is generally recognised as an urgent mitigation measure (Rubin et al., 2013). The amine-based post-combustion CO₂ capture technology is the most matured and promising technology option (Nwaoha, 2018). However, its industrial deployment requires large investments as well as enormous energy supply for desorption (Lim et al., 2013; Aromada and Øi, 2017). Therefore, the need for efficient cost and economic analysis aimed at reduced CO₂ capture investment and operating costs is imperative.

The lean/rich heat exchanger is one of the most expensive equipment in an amine-based CO₂ capture plant, and it has a considerable cost implication on the investment (Ali et al., 2019).

In preliminary cost estimation of heat exchangers, the important design parameter is the heat transfer area needed. That is evaluated from the heat duty (heat transfer from hot to cold stream), overall heat transfer coefficient, and the log-mean temperature difference (LMTD) (van der Spek et al., 2019). However, the required heat duty depends on the minimum approach temperature (ΔT_{min}).

In post-combustion solvent-based CO₂ capture, studies on cost optimisation of the lean/rich MEA heat exchanger have been based on ΔT_{min} of the shell and tube heat exchanger (STHX) types (Kallevik, 2010; Øi et al., 2014; Aromada and Øi, 2017; Ali et al., 2019). None of such studies has been found for the plate and frame heat exchanger (PHE). Thus, this study is conducted on cost optimisation of the PHE based on ΔT_{min}. This is carried out by performing process simulations of CO₂ absorption and desorption process. Cost estimation and optimisation to find the most cost effective and technically suitable type of heat exchanger for the lean/rich heat exchanger is then carried out.

1.1 Process Description and Scope

The process comprises a flue gas fan for transporting the flue gas through the direct contact cooler (DCC) where the temperature is reduced. The DCC pump and DCC cooler help in circulation and cooling of the water respectively. The main capture process consists of an absorber, a desorber with a reboiler at the bottom and a condenser, lean/rich heat exchanger, pumps and a cooler. *Figure 1* shows the flowsheet of the standard capture process.
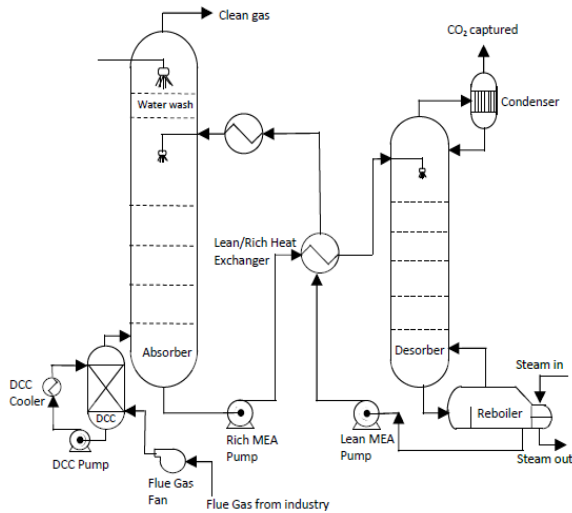
Figure 1. . Flowsheet of the standard process

2 Simulation, Specifications and Assumptions

2.1 Specifications for Simulation

Table 1 presents the specifications used for the base case simulations. The flue gas data are from a cement industry and are taken from (Onarheim et al., 2015; Ali et al., 2019).

2.2 Process Simulation

Aspen HYSYS Version 10 is used for the simulations with the same calculation approach as in (Øi, 2007; Aromada and Øi, 2015). The difference is that in version 10, the acid gas

property package replaces the Amine property package in previous versions.

Table 1. Specifications for simulation (Onarheim et al., 2015; Ali et al., 2019)

Specifications	
Flue gas	
Temperature [°C]	80
Pressure [kPa]	121
CO_2 mole-fraction	0.2520
H_2O mole-fraction	0.0910
N_2 mole-fraction	0.5865
O_2 mole-fraction	0.0705
Molar flow rate [kmol/h]	11472
Flue gas from from DCC to absorber	
Temperature [°C]	40
Pressure [kPa]	121
Lean MEA	
Temperature	40
Pressure [kPa]	121
Molar flow rate [kmol/h]	96850
Mass fraction of MEA [%]	29
Mass fraction of CO_2 [%]	5.30
Absorber	
No. of absorber stages	15
Absorber Murphree efficiency [%]	11- 21
ΔT_{min}, lean/rich heat exchanger [°C]	10
Desorber	
Number of stages	10
Desorber Murphree efficiency [%]	100
Pressure [kPa]	200
Reboiler temperature [°C]	120
Reflux ratio in the desorber	0.3
Temperature into desorber [°C]	104.6
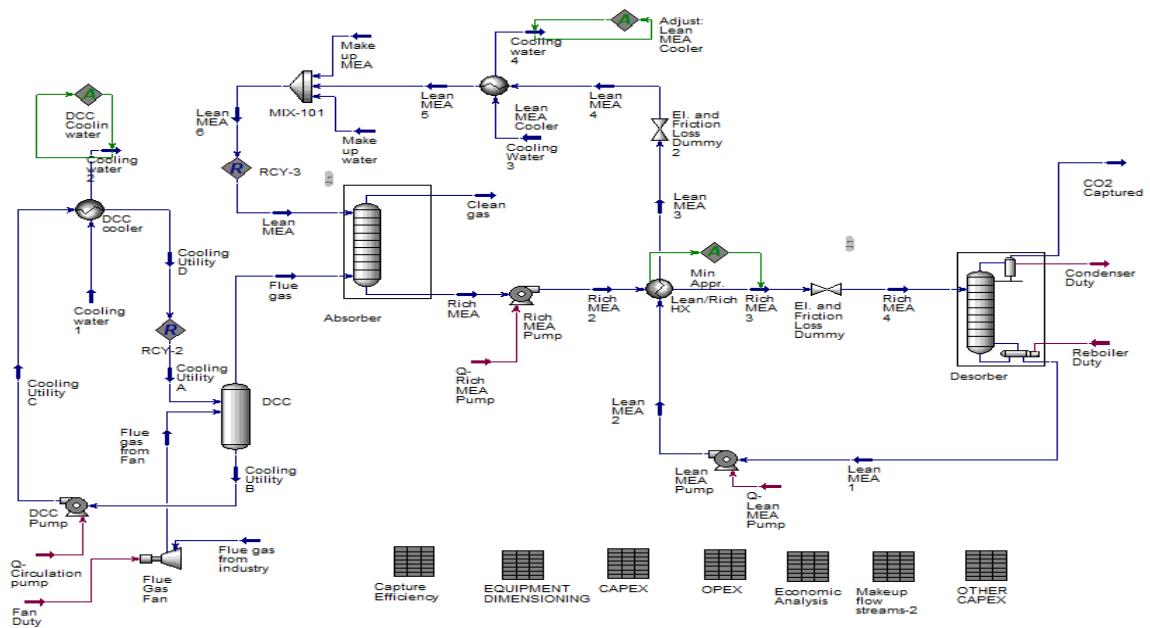
Figure 2. Aspen HYSYS flowsheet

The absorption column as well as the desorption columns are both simulated as equilibrium stages with stage efficiencies. The absorber is simulated with 15 packing stages, while it is 10 packing stages for the desorber. Murphree efficiencies for CO_2 are specified in the simulation. For more details on Murphree efficiency, see (Øi, 2007). Equilibrium stages of 1 m height each for both columns are assumed. Murphree efficiencies of 11 – 21% were specified from bottom to the top of the absorption column (Ali et al., 2019). A constant Murphree efficiency of 100% is specified for all the stages of the desorption column. The Modified HYSIM Inside-Out algorithm was selected in the columns because it helps to improve convergence (Aromada and Øi, 2015).

Adiabatic efficiency of 75% was specified for all the pumps and the flue gas fan. The Aspen HYSYS simulation process flow diagram (PFD) is given in *Figure 2*.

3 Methods

3.1 Scope of the Cost Estimates

The equipment included in this cost analysis are for cooling the flue gas before it enters the absorption column, and for the absorption and desorption process as can be seen in *Figure 1* and *Figure 2*. The study does not include equipment for pre-treatment unit of the flue gas and water-wash section. The equipment for CO_2 compression are not considered because the focus is on the lean/rich heat exchanger

The total investment cost in this study is limited to the sum of the installed costs of the equipment considered. The cost of acquiring the site (land), preparing the site and for service buildings are not included.

The operating and maintenance costs (OPEX) include the cost of electricity, steam, cooling and process water, solvent (MEA), salaries of 6 operators and 1 engineer, and annual maintenance cost set at 4 % of the installed cost of the equipment as given in *Table 2*.

Table 2. Operating cost data

	Unit	Value/unit*	Reference
Steam	€/kWh	0.032	Husebye et al. (2012)
Electricity	€/kWh	0.132	Ali et al. (2019)
Water	€/m³	0.022	Ali et al. (2019)
MEA	€/m³	2 069	Ali et al. (2019)
Maintenance	€	4% of CAPEX	Ali et al. (2019)
Operator	€	85 350 (x6)	Ali et al. (2019)
Engineer	€	166 400	Ali et al. (2019)

The costs have been escalated to January 2020

Costs for CO_2 transport and storage, pre-production costs, insurance, taxes, first fill cost and administrative costs are not included in the OPEX.

3.2 Equipment Dimensioning and Assumptions

Dimensioning of equipment in this study follows the approach used in Ali *et al.* (2019) based on mass conservation and energy balances of the system. *Table 3* summarises the dimensioning factors and assumptions used in this work.

Table 3. Equipment dimensioning factors and assumptions

Equipment	Sizing factors	Basis/Assumptions
DCC Unit	Tangent-to-tangent height (TT), Packing height, internal and external diameters (all in [m])	Velocity using Souders-Brown equation with a k-factor of 0.15 m/s (Yu, 2014, pp. 97). TT =15 m, 1 m packing height/stage (4 stages)
Absorber		Superficial velocity of 2 m/s, TT=40 m, 1 m packing height/stage (15 stages)
Desorber		Superficial velocity of 2 m/s, TT=22 m, 1 m packing height/stage (10 stages)
Lean/rich heat exchanger	Heat transfer area, A [m²]	U = 0.5 kW/m².K (Ali et al., 2019)
Reboiler		U = 0.8 kW/m².K (Ali et al., 2019)
Condenser		U = 1.0 kW/m².K (Ali et al., 2019)
Coolers		U = 0.8 kW/m².K (Ali et al., 2019)
Pumps	Flow rate [l/s]	Centrifugal
Flue gas fans	Flow rate [m³/h]	Centrifugal

3.3 Cost Estimation and Assumptions

The Enhanced Detailed Factor (EDF) method is used for estimation of all the equipment costs and overall plant investment cost. Readers are referred to Ali *et al.* (2019) for the details and application of the EDF method.

The purchased costs of the equipment are obtained from Aspen In-plant Cost Estimator Version 11 with a cost year of 2018 (January). The costs are then escalated to January 2020 using the SSB (Norwegian Statistisk sentralbyrå, webpage)

industrial cost index (2018 = 106; 2020 = 111.3). The currency conversion rate for Euro to NOK is 10.13, taken from (NorgesBank, 2020 webpage) on January 25, 2020. Conversion to NOK is necessary to use the enhanced factors developed by Nils Eldrup (Ali et al., 2019). The default location is Rotterdam in Netherlands.

All equipment is assumed to be made from stainless steel (SS316), except the Flue gas fan, which is from carbon steel (CS). Material factor to convert costs in SS316 to CS is 1.75 and 1.30 for seamless and welded equipment respectively.

A brownfield, and an Nth-of-a-kind (NOAK) project are assumed. 25 years of project, of which 2 years are for plant construction, and 7.5% interest rate is also assumed (Ali et al., 2019).

4 Results and Discussion

4.1 Simulation Results

Table 4 presents the process simulation results for the base case and sensitivity analysis of ΔT_{min}. Lower ΔT_{min} give lower reboiler heat and lower lean MEA cooler duty (more heat has been transferred from the lean stream to the rich stream). Therefore, less steam and less cooling water are required in the reboiler and lean MEA cooler respectively.

Table 4. Simulation results

ΔT_{min}	Reboiler heat	Typical results	Lean MEA cooler duty
[°C]	[GJ/ton CO_2]		[kW]
5	3.83		66 389
10	4.08	3.3 - 5.0	81 333
15	4.27	(Nwaoha	89 333
20	4.67	et al.,2018)	117 778

4.2 Base Case Plant Investment Cost

The base case in this study has a U-tube shell and tube heat exchanger. The total investment cost (CAPEX) which is the sum of the installed costs of all the equipment is €97.5 million. The cost estimation results obtained show the same trends with similar studies by Ali et al. (2018) and Ali et al. (2019). The lean/rich heat exchanger contributes most to the total investment cost compared to other equipment as in *Figure 3*.

The heat exchanger accounts for 41% of the total capital cost (*Figure 4*). Ali et al. (2019) also calculated the lean/rich heat exchanger to have the highest installed cost for the same scope as in this study. It accounts for 37% of the CO_2 capture plant. They obtained their cost data from Aspen In-plant Cost Estimator V10 with a cost data year of 2016.

Aspen In-plant Cost Estimator V11 with a cost data year 2018 is used in this study. In addition, the cost in this study are escalated from 2018 to 2020. This explain the 3% difference from a similar process. In the work of Nwaoha *et al.* (2018), for a process with an absorber packing height and diameter of 21.95 m and 10.07 m respectively, the lean/rich heat exchanger has the second highest cost for both MEA and AMP-PZ-MEA systems. The absorber in their case has the highest cost. The diameter is almost twice and the packing height is approximately 7 m higher than in this work. The study was for a 90% CO_2 capture process from a cement plant flue gas with 0.115 mole of CO_2. In this study, capture efficiency is 85% and the CO_2 molar composition is 0.252.
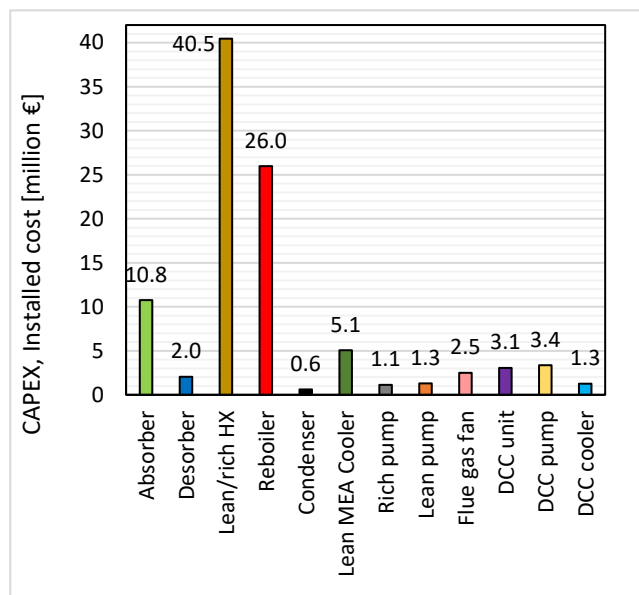
Figure 3. Equipment installed costs of the base case 85% CO_2 capture plant

4.3 Operation and Maintenance Costs

The annual operation and maintenance (O&M) cost for the base case is €44.5 million. Only the steam consumption costs €31.7 million and annual maintenance cost is €3.9 million.

4.4 Annualised CAPEX, Total Annual Cost and Capture Cost

Annualised capital cost is obtained from the following relation:

$$Annualised\ CAPEX = \frac{CAPEX}{Annualised\ factor} \qquad (1)$$

The annualised factor is calculated as follows:

$$Annualised\ factor\ = \ \sum_{i=1}^{23}\left[\frac{1}{(1+r)^n}\right] \quad (2)$$

Where *n* represents operational years and *r* is discount/interest rate. The annualised CAPEX for the base case is evaluated to be €9 million (CO_2 compression equipment not included). Thus, the total annual cost, which is the sum of the annualised CAPEX and the yearly OPEX, is €53.6 million. Figures 4 presents the annual cost distribution. The CO_2 capture cost is estimated from:

$$CO_2\ capture\ cost = \frac{Total\ annual\ cost}{Mass\ of\ CO_2\ Captured} \quad (3)$$

The CO_2 capture cost for the base case is 61.9 €/ton CO_2 (2020). In the literature, it is between €50/ton CO_2 – 128/ton CO_2 (Ali et al., 2019). (Ali et al., 2019) calculated this cost for a similar process but with the compression section to be €62.5/ton CO_2 for a cost year of 2016. For a full process that include compression, Nwaoha et al. (2018) calculated this cost for 90% CO_2 capture from a cement plant flue gas with CO_2 compression to be US$93.2/ton CO_2 (i.e., €74.5/ton CO_2). According to Irlam (2017), for a first-of-a-kind (FOAK) CSS complete technology, the CO_2 avoided cost for the cement industry is US$188 (€164.4) and US$130 (€113.7) per ton CO_2 for Germany and Poland respectively. FOAK technologies usually cost between 15 – 55% more than NOAK (Boldon & Sabharwall, 2014).
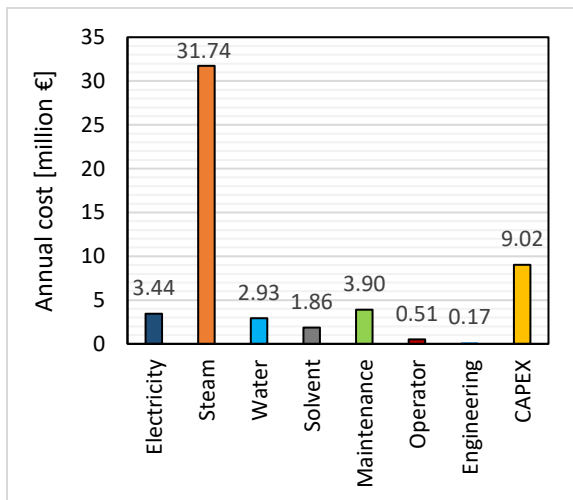
Figure 4. Cost distribution of the base case total annual cost

(Carbon Capture & Storage Association, 2011-2020) states that the capture cost range is €60/ton CO_2 – €90/ton CO_2 for the power industry. They projected that it will reduce considerably to €35 – 50/ton CO_2 in the beginning of 2020. Based on *Figure 4*, this reduction will have to come from reducing mainly the cost of steam. This can be achieved using available waste heat to generate

steam or very cheap steam for desorption (Ali et al., 2018). Electricity cost is low in this study compared to Nwaoha et al. (2018) and Ali et al. (2019). This is because CO_2 compression is not considered in this work. The compressors require much more electrical energy compared to pumps and fan/blower.

4.5 CAPEX Based on Different Heat Exchangers

Figure 5 presents the total installed cost of CO_2 capture plant options of using the different types of heat exchangers. The compact heat exchangers offer considerable lower total investment cost compared to the conventional shell and tube heat exchangers. Using the gasketed-plate heat exchanger (G-PHE) will give the lowest plant investment cost. The purchase cost of the welded-plate heat exchanger (W-PHE) was assumed to be 30 % more expensive than the G-PHE based on information from Peters et al. (2004).

The reference case, which has U-tube shell and tube heat exchanger (UT-STHX) has investment cost of €97.5 million. The case with fixed tube sheet heat exchanger (FTS-STHX) has a CAPEX of €102.4 million. The installed cost of the plant with G-PHE is €72.6 million. The plant option with floating-head shell and tube heat exchanger (FH-STHX) gives the highest installed cost of €103.8 million.
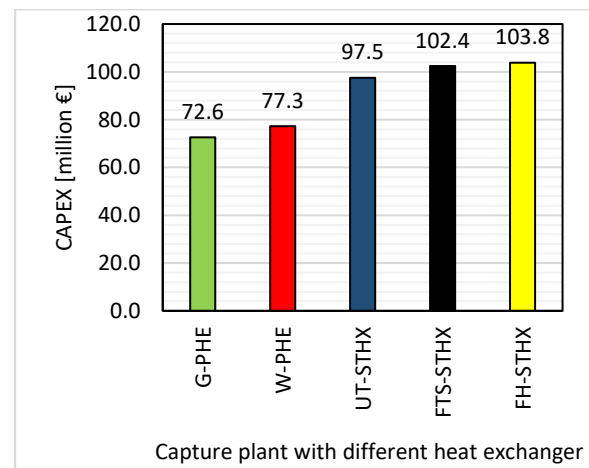
Figure 5. Total plant installed costs for different heat exchangers

4.6 Optimisation: Minimum Approach Temperature

Cost optimisation of the lean/rich heat exchanger in this study is done by finding the cost optimum minimum approach temperature (ΔT_{min}).

The plants with G-PHE and welded-plate heat exchanger (W-PHE) have their minimum CAPEX

at 15℃, while it is 20℃ for the 4 STHXs. As the ΔT_{min} increases, the heat transfer area is reduced, thereby reducing the CAPEX since the lean/rich heat exchanger with STHXs account for 41 – 45 % of the CAPEX in this study. The slight increase of CAPEX from 15 – 20℃ as can be observed in Figure 6 for the PHEs is caused by increase in the cost of other equipment like the lean MEA cooler and the reboiler. This will also result in higher OPEX, especially from higher steam consumption as can be seen in *Table 3* and *Figure 7*. More cooling water is also needed. However, increase in OPEX is slight from 5 – 15℃ for the STHXs but becomes significantly steep from 15 – 20℃. That is the same for the PHEs except that the OPEX is considerably lower at 5℃ compared to 10℃.

In order to find the optimum design ΔT_{min} we evaluated the CO_2 capture cost at the different ΔT_{min} for the different heat exchanger options. **Figure 8** presents the results.

The STHXs and W-PHE have their optimum CO_2 capture costs at 15℃. While the G-PHE optimum cost is at 5℃, which is due to its relative lower cost per heat transfer area and lower maintenance cost. Cost savings of €1.6/tCO₂, €1.1/tCO₂ and €1.0/tCO₂ are achieved by the cost optimum cases with U-tube, fixed tube-sheets and floating-head shell and tube heat exchangers when compared with the base case. The cost optimum cases with gasketed and welded plate heat exchangers have a cost savings of €4.0/tCO₂ and €3.4/tCO₂ respectively.
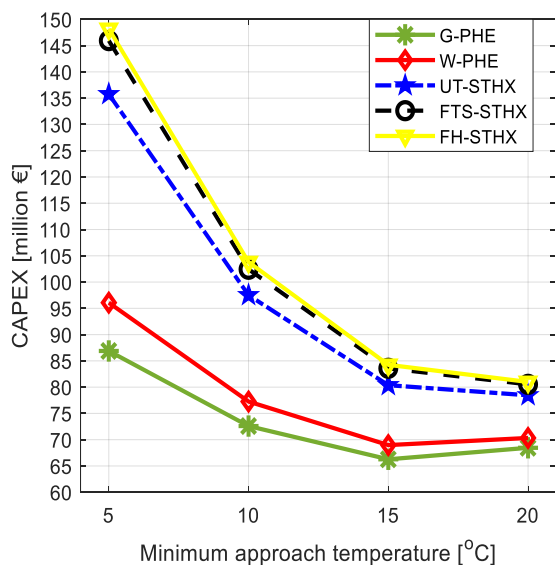
Figure 6. CAPEX of the different heat exchangers at different ΔT_{min}

All the studies of optimum ΔT_{min} we found of solvent-based CO_2 capture used STHXs (Kallevik, 2010; Øi et al. (2014), Li et al., 2016; Aromada & Øi, 2017; Nwaoha et al., 2018; Ali et al., 2019). None of such studies was found for other types of

heat exchangers like PHE. This is because in the chemical industry, about 60% of heat exchangers in use are STHX (Peters et al., 2004). They are more robust, they can be applied in all types of processes, they can withstand higher pressures, higher temperatures and thermal stresses, and higher pressure difference between the hot and cold streams.

Additional advantage of the STHX is that they have well-established design codes, standards and specifications, especially by TEMA (Tubular Exchanger Manufacturers Association) and American Society for Mechanical Engineers (ASME). The PHEs do not have such established or accepted design standards. Therefore, higher design uncertainties are expected for the PHEs.
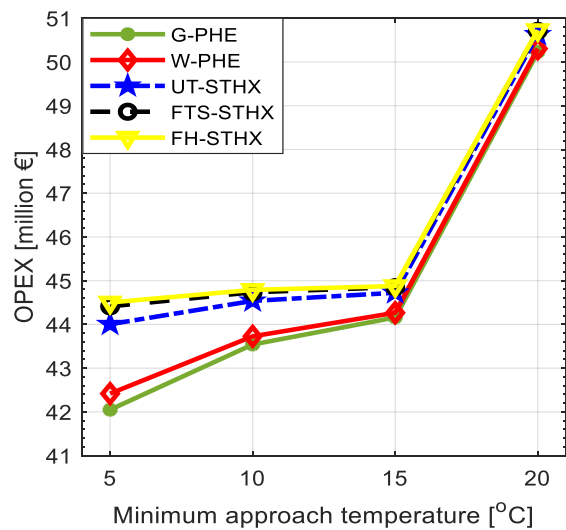
Figure 7. OPEX of the different heat exchangers at different ΔT_{min}
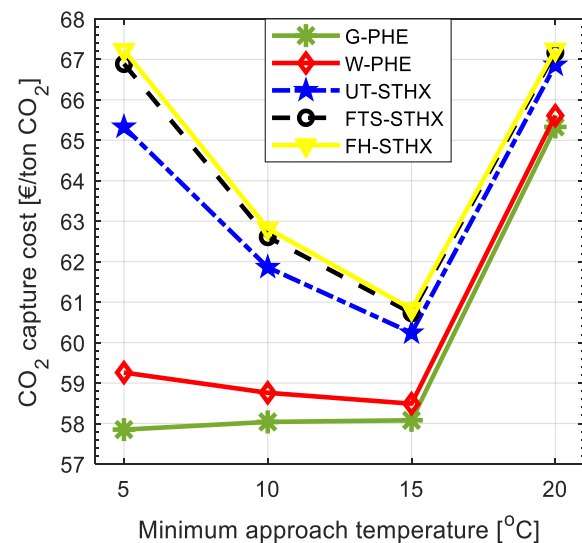
Figure 8. Cost optimum ΔT_{min} of the different heat exchangers

Nonetheless, the plate and frame heat exchangers are increasingly being considered for

application in the process industry (Peters et al., 2004). This is because capital-intensive processes need heat exchangers that can achieve higher thermal efficiencies and simultaneously reducing equipment/investment costs (Peters et al., 2004). The PHEs also occupy less space and have less weight for the same heat transfer area as STHX.

4.7 Maintenance

Maintenance of the G-PHE is easier and far less expensive. The plates are removable and can easily be cleaned mechanically. Thus, it is the most ecological option (Marcano, 2015). The parts can easily be replaced relatively inexpensively.

The plates of the W-PHE are welded and thus are not removable. Consequently, cleaning can only be done by chemical means. The only advantage over G-PHE is that W-PHE can withstand higher pressures and temperatures. This advantage is not relevant in CO_2 solvent-based absorption and desorption systems since the pressures and temperatures are relatively low

FTS-STHX and FH-STHX are cleaned both mechanically (inside the tubes) and chemically (outside surfaces of tube). The UT-STHX normally requires only chemical cleaning because of the U-tube shape of the tubes.

Therefore, the G-PHE which require less space is the most ecologically friendly option, and it is the easiest and the cheapest to maintain among the heat exchangers investigated.

4.8 Maintenance and Operating Cost Discussion

Figure 7 presents the OPEX calculated in this study. At ΔT_{min} of $5 - 10°C$, the calculated annual OPEX of the PHEs is considerably less than that of any of the STHXs, even though the difference in this study is only based on maintenance cost. The gap gets closer at $15°C$ and and very close at $20°C$. This is because the purchase and installed costs of the STHXs reduce drastically at ΔT_{min} 15 and $20°C$.

4.9 Comparison with Previous Studies

In this section, comparison of this study is done with some previous studies. All the previous studies used the STHX.

(Øi et al., 2014) calculated the cost optimum ΔT_{min} of a plant with 16 absorber packing stages, and 20 years period with discount rate of 10.5% to be 12°C.

(Aromada and Øi, 2017) estimated it to be 13°C for a system with 15 absorber packing stages with discount rate of 7% for 20 operational years, based on negative-NPV method. When the years of plant

operation were reduced to 15 years, cost optimum ΔT_{min} became 14°C.

(Kallevik, 2010), also applied negative NPV for 20 years calculation period, with 7% discount rate, estimated the cost optimum ΔT_{min} to fall within 10 – 14°C, for a 85% CO_2 capture with 15 absorber packing stages.

Most recent is (Ali et al., 2019), for a calculation period of 24 years and interest rate of 7.5%, evaluated the cost optimum ΔT_{min} to be 10°C.

These results suggest that the cost optimum ΔT_{min} for the STHXs is within $10 - 16°C$, which are in agreement with this study. The little differences obtained from the different studies occur due to the different sources of cost data and economic assumptions like interest rates and operational years.

Several technical studies have also shown that operating at 5°C ΔT_{min} will help in reduction of the reboiler heat in CO_2 capture processes. However, the capital cost of achieving this makes it not to be the cost optimum design parameter for the well-established STHX. This study suggests that ΔT_{min} of 5°C or between $5 - 10°C$ can be energy optimum and cost optimum design if G-PHE is used.

5 Conclusion

Simulations of 85% CO_2 absorption and desorption process aimed at cost optimisation of the lean/rich heat exchanger has been performed using Aspen HYSYS Version 10. This was followed by cost estimation and optimisation of the lean/rich heat exchanger by finding the type of heat exchanger and the design optimum ΔT_{min} among 5, 10, 15 and 20°C ΔT_{min}. Considerable savings in capital and operating costs can be achieved by selecting the plate and frame heat exchanger instead of the conventional shell and tube types, in a CO_2 absorption and desorption plant design. The PHEs require only 30, 15, 9, and 6 number of units for the cases of 5, 10, 15 and 20°C ΔT_{min} respectively, compared to 44, 23, 13, 9 number of units respectively for the STHXs. The G-PHE gives the lowest total annual cost in all the ΔT_{min}. G-PHE with 5°C ΔT_{min} is calculated to be the energy optimum and the cost optimum design for the lean/rich heat exchanger.

Abbreviations

PHE: Plate and frame heat exchanger
G-PHE: Gasketed-plate or plate and frame heat exchanger
W-PHE: Welded- plate heat exchanger
STHX: Shell and tube heat exchanger
UT-STHX: U-tube shell and tube heat exchanger
FTS-STHX: Fixed-tube sheet Shell and tube heat exchanger
FH-STHX: Floating head shell and tube heat exchanger

References

H. Ali, N.H. Eldrup, F. Normann, R. Skagestad, and L. E. Øi. Cost Estimation of CO_2 Absorption Plants for CO_2 Mitigation–Method and Assumptions. *International Journal of Greenhouse Gas Control*, 88, 10-23, 2019. doi: 10.1016/j.ijggc.2019.05.028

H. Ali, L. E. Øi, and N. H. Eldrup. Simulation and Economic Optimization of Amine-based CO_2 Capture using Excess Heat at a Cement Plant *Linköping University Electronic Press Conference Proceedings (SIMS 59)*, 153:58-64, 2018. doi:10.3384/ecp1815358

S. A. Aromada and L. E. Øi. Simulation of Improved Absorption Configurations for CO_2 Capture. *Linköping Electronic Press Conference Proceedings (SIMS 56)*, 119(2):21-29, 2015. doi: 10.3384/ecp1511921

S. A. Aromada and L. E. Øi. Energy and Economic Analysis of Improved Absorption Configurations for CO_2 Capture. *Energy Procedia*, 114: 1342-1351, 2017. doi: 10.1016/j.egypro.2017.03.1900

L. M. Boldon and P. Sabharwall. *Small modular reactor: First-of-a-Kind (FOAK) and Nth-of-a-Kind (NOAK) Economic Analysis* (No. INL/EXT-14-32616). Idaho National Lab. (INL), Idaho Falls, ID (United States), 2014. doi: 10.2172/1167545

Carbon Capture and Storage Association (Webpage). *Affordability, CCS: Keeping the lights on without costing the earth.* 2011 – 2020. Available: http://www.ccsassociation.org/why-ccs/affordability/ Accessed 07.05.2020

J. Husebye, A. Brunsvold, S. Roussanaly and X. Zhang. Techno economic evaluation of amine based CO_2 capture: impact of CO_2 concentration and steam supply. *Energy Procedia*, 23, 381 – 390, 2012. doi: 10.1016/j.egypro.2012.06.053

L. Irlam. *Global costs of carbon capture and storage*. Global CCS Institute, Melbourne, Australia, 2017.

O. B. Kallevik. *Cost estimation of CO_2 removal in HYSYS*. Master's thesis. Høgskolen i Telemark, 2010.

Y. Lim, J. Kim, J. Jung, C. S. Lee, and C. Han. Modeling and simulation of CO_2 capture process for coal-based power plant using amine solvent in South Korea. *Energy Procedia*, 37:1855-1862, 2013. doi.org/10.1016/j.egypro.2013.06.065

L. A. Marcano. *Design and evaluation of heat exchangers used in post-combustion CO_2 capture plants*. Master's Thesis. Telemark University College, Porsgrunn, 2015.

Norgesbank webpage. Available on: https://www.norges-bank.no Accessed: 25.01.2020.

C. Nwaoha, M. Beaulieu, P. Tontiwachwuthikul and M. D. Gibson. Techno-economic analysis of CO_2 capture from a 1.2 million MTPA cement plant using AMP-PZ-MEA blend. *International Journal of Greenhouse Gas Control*, 78:400-412, 2018. doi.org/10.1016/j.ijggc.2018.07.015

K. Onarheim, S.Ò. Garðarsdòttir, A. Mathisen, L. O. Nord and D. Berstad. *Industrial implementation of carbon capture in Nordic industry sectors*. Nordic CCS Competence Centre NORDICCS 2015.

M. S. Peters, K. D. Timmerhaus, and R. E. West. *Plant Design and Economics for Chemical Engineers*, 5th edition, McGraw-Hill Companies, Inc. Singapore 2004.

E. S. Rubin, C. Short, G. Booras, J. Davison, C. Ekstrom, M. Matuszewski and S. McCoy. A proposed methodology for CO_2 capture and storage cost estimates. *International Journal of Greenhouse Gas Control*, 17, 488-503, 2013 doi: 10.1016/j.ijggc.2013.06.004

SSB (Norwegian Statistisk sentralbyrå) webpage. Industrial cost index. Available on: https://www.ssb.no/en Accessed 25.01.2020.

M. van der Spek, S. Roussanaly and E. S: Rubin. Best practices and recent advances in CCS cost engineering and economic analysis. *International Journal of Greenhouse Gas Control*, 83:91-104, 2019. doi: 10.1016/j.ijggc.2019.02.006

F. Yu. *Process design for chemical engineers*. Ten Books, Inc., 2014.

L. E. Øi. Aspen HYSYS simulation of CO2 removal by amine absorption in a gas based power plant. The 48th Scandinavian Conference on Simulation and Modelling (SIMS 2007), Göteborg, Sweden, *Linköping University Electronic Press*, 27:73-81, 2007.

L. E. Øi, T. Bråthen, C. Berg, S. K. Brekne, M. Flatin, R. Johnsen, I. G. Moen, and E. Thomassen. Optimization of configurations for amine based CO_2 absorption using Aspen HYSYS. *Energy Procedia*, 51:224-233, 2014. doi: 10.1016/j.egypro.2014.07.026

Process Simulation, Cost Estimation and Optimization of CO$_2$ Capture using Aspen HYSYS

Lars Erik Øi[*] Nils Eldrup Solomon Aromada Andrea Haukås Joakim Helvig
Ida Hæstad Anne Marie Lande

Department of Process, Energy and Environmental Technology, University of South-Eastern Norway
Corresponding author's email address: lars.oi@usn.no

Abstract

A standard method to remove CO$_2$ is by absorption in monoethanol amine (MEA) followed by desorption. A traditional aim has been to find the process parameters which give the lowest combined investment and operating cost. The aim in this work is to calculate cost optimum process parameters and evaluate whether it is possible to perform automated cost estimation and optimization. Aspen HYSYS simulations of a standard amine based process for CO$_2$ capture from a cement plant have been performed. The capital cost of CO$_2$ capture was estimated based on equipment cost from Aspen In-plant cost estimator and a detailed factor method. Operating cost included electricity, heat consumption and maintenance. Optimum temperature difference in the main heat exchanger was calculated to 13 °C after one simulation for each temperature. The lowest calculated cost was achieved with 12 stages (meter packing height) based on one simulation for each stage number. With improved robustness of the simulations, it should be possible to optimize the temperature difference in one automated calculation. To optimize the height of the absorption colum^n automatically, a way to update the number of stages during the simulations has to be found.

Keywords: carbon capture, Aspen HYSYS, simulation, cost estimation

1 Introduction

The cement industry accounts for more than 5 % of the total anthropogenic emissions of CO$_2$ in the world today (Norcem, 2019). There are several possible options to reduce emissions in the cement industry, one of them is CCUS (carbon capture, utilization and storage). Absorption using amine solutions is considered the most favorable method for capture of CO$_2$ from exhaust gas.

This work is based on the project work from Haukås et al. (2019) at the University of South-Eastern Norway (USN). It is a continuation of previous work at Telemark University College (TUC) and USN. This work has involved process simulation, equipment dimensioning, cost estimation and cost optimization of CO$_2$ capture. The simulation tool Aspen HYSYS has been used in most of the work, with the application of the amine package and constant stage (Murphree) efficiencies in the absorber and desorber.

1.1 Aim

The general aim of this project is to develop further models in Aspen HYSYS for calculation, equipment dimensioning, cost estimation and optimization of CO$_2$ capture by atmospheric exhaust gas absorption into an amine solution. The intention is to streamline the cost estimation and optimization procedure by utilizing the spreadsheet function in Aspen HYSYS. One specific aim is to calculate cost optimum process parameters and evaluate whether it is possible to perform automated cost estimation and optimization.

1.2 Literature

The combination of process simulation and cost estimation is an important tool to evaluate different CO$_2$ capture technologies (Rao and Rubin, 2002; Øi, 2012; Ali, 2019).

This work is a continuation of previous work of students at TUC and USN. In particular, the work is based on the master project from 2015 (Park et al., 2015), as well as the master thesis (Kallevik, 2010).

The project from 2015 involved process simulation, dimensioning and cost estimation of CO$_2$ capture from a cement plant with the use of Aspen HYSYS. MEA was the sorbent, and the amine package in Aspen HYSYS was applied. Capture rate, energy demand per kg CO$_2$ captured, and capture cost per ton CO$_2$ were calculated. The base case had a CO$_2$ removal efficiency of 90 %. The Aspen HYSYS simulation performed in 2015 did not have any adjust operations incorporated.

In the master's thesis by Kallevik (2010), a simulation and cost estimation of a carbon capture process was developed in Aspen HYSYS in an attempt to cost optimize the process. By varying the removal efficiency and changing process parameters, such as minimum temperature difference in the main heat exchanger, an optimum solution for the minimum temperature difference was found.

The thesis by Kallevik used the amine package for his simulations in Aspen HYSYS. In addition, an adjust operation was implemented in the flowsheet in order to get an automated model to specify the minimum approach temperature in the heat exchanger.

1.3 Process description

Figure 1 shows a standard process for CO_2 absorption into an amine-based solvent. It comprises an absorption column, a stripping column including a reboiler and condenser, circulating pumps and heat exchangers. The process is described in more detail in Kallevik (2010) and Øi (2012).
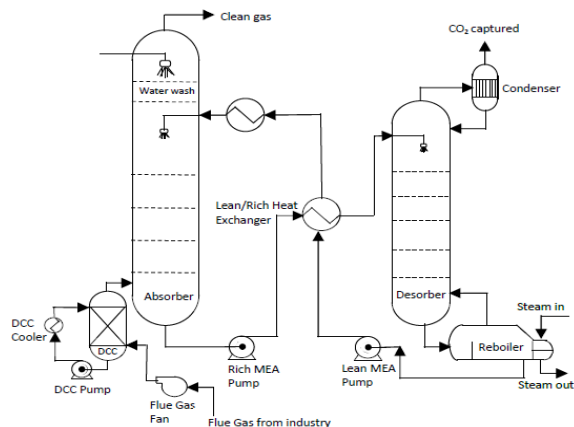
Figure 1. Process flow diagram of a standard amine-based CO_2 capture process

2 Methodology

The process simulation tool Aspen HYSYS version 10 was used for all the simulations performed in this report. The amine package (which is now the recommended equilibrium model by Aspen HYSYS) and constant Murphree efficiencies were specified in the absorber and desorber.

2.1 Specifications and simulation of standard CO₂ capture process

The specifications in Table 1 correspond to 90 % CO_2 removal efficiency and a minimum approach temperature of 10 °C in the lean/rich heat exchanger. This is the base case alternative.

The calculation sequence is similar to earlier works (Aromada and Øi, 2015). Even though Aspen HYSYS is an equation based program, the calculation strategy is based on a sequential modular approach (Kisala et al., 1987; Ishii and Otto, 2008). First the absorption column is calculated from the inlet gas and the lean amine (which is first guessed). The rich amine from the bottom of the absorption column passes through the pump and the rich/lean heat exchanger. The temperature after the heat exchanger is specified. The heated rich amine is entering the desorption column which calculates the CO_2 product and the hot lean amine. The hot lean amine is pumped to a higher pressure in a pump, passes through the lean/rich heat exchanger and is further cooled in the lean cooler. Then this lean amine is

checked in a recycle block. It is checked whether the flow and composition in the recycled lean amine is sufficiently close to the earlier guessed lean amine stream, which may be changed by iteration.

Two adjust operations were implemented in the flowsheet in order to get an automated simulation model. One is adjusting the minimum approach temperature in the lean/rich heat exchanger and another adjusting the removal efficiency based on the lean amine mass flow. The process flowsheet is shown in Figure 2.

Table 1. Aspen model parameters and specifications for the base case alternative

Parameter	
Inlet flue gas temperature [°C]	40.0
Inlet flue gas pressure [kPa]	110.0
Inlet flue gas flow rate [kmol/h]	8974
CO_2 content in inlet gas [mole %]	17.8
Water content in inlet gas [mole %]	19.5
Lean amine temperature [°C]	45.0
Lean amine pressure [kPa]	101.0
Lean amine rate [kg/h]	$1.103 \cdot 10^6$
MEA content in lean amine [mass %]	28.71
CO_2 content in lean amine [mass %]	5.16
Number of stages in absorber [-]	10
Murphree efficiency in absorber [m⁻¹]	0.15
Rich amine pump pressure [kPa]	220.0
Rich amine temp. out of HEX [°C]	102.8
Number of stages in desorber [-]	6
Murphree efficiency in desorber [m⁻¹]	1.0
Reflux ratio in stripper [-]	0.3
Reboiler temperature [°C]	120.0
Lean amine pump pressure [kPa]	200.0

2.2 Parameter variations

Three different parameters were varied in order to study the effects on the cost estimate:

1. Minimum approach temperature in lean/rich heat exchanger
2. CO_2 removal efficiency
3. Number of stages in the absorber

10 stages, removal efficiency of 90 % and a 10 °C minimum approach temperature correspond to the base case simulation. In the additional cases, all the base case parameters were constant except the parameter to be optimized.

2.3 Process convergence and tolerances

To converge a column model in a simulation software tool, e.g. Aspen HYSYS, all equations describing equilibrium, gas and liquid flow must be solved for each calculation stage. Aspen HYSYS has a default set of

criterias for converging of column models, and also preset calculation parameters. References for flow-sheet convergence and stage-to-stage column convergence are Kisala et al. (1987), Ishii and Otto (2008) and Holoboff (2020).

In order to achieve convergence, there are different calculation models available in Aspen HYSYS. In the columns, the best convergence is achieved by using the Modified Hysim Inside-Out algorithm with adaptive damping in absorber and desorber (Øi, 2012).

Aspen HYSYS is an equation-based simulation software, which means that it has the ability to calculate in-streams based on out-streams. However, the calculation strategy in this work is sequential. In the case of recycle streams, one must include recycle blocks to solve the flowsheet in Aspen HYSYS. This block compares the in-stream to the block with the out-stream from the block with the previous iteration.

The goal of tolerance testing is to reach a specified target value with an accuracy up to a certain number of decimals. An investigation of the column and flowsheet convergence in the base case simulation was performed by decreasing the tolerances of the adjust operations. The secant method in Aspen HYSYS was applied for all alternatives.

In order to reach the target value of minimum approach temperature equal to 10 °C, the temperature of rich amine out of the heat exchanger was varied with the adjust operation. Each new trial had a more restrictive tolerance than the previous.

The varied parameter in the testing of the removal efficiency is the mass flow of the lean amine into the absorber. The target value is a cleaning efficiency of 90.00 % that is calculated from the molar flow of CO_2 in the cleaned gas out of the absorber.

The tolerance of the recycle operation was also decreased. In the recycle operation the tolerance of various parameters like temperature, pressure, flow, composition and individual components can be modified independently.

2.4 Dimensioning and cost estimation calculations

The following procedure was implemented for the cost estimation:

1. Simulation of the CO_2 capture process in Aspen HYSYS with the base case specifications in Table 1.
2. Dimensioning of the equipment based on the simulation result
3. Calculation of equipment cost for each unit using Aspen In-Plant cost estimator
4. Calculation of installation cost based on a detailed factor table. The factor was kept constant under parameter variation
5. Correction of currency and index

6. Estimation of annual operational costs based on energy requirement from the simulation result
7. Calculation of net present value based on a given discount rate and project lifetime
8. Calculation of CO_2 capture cost for comparison between the different case simulations

2.4.1 Scope analysis

The cost analysis is limited to the equipment listed in the flow-sheet in Figure 1 excluding the flue gas cooler. No pre-treatment like inlet gas purification or cooling is considered, and no treatment after stripping like compression, transport or storage of CO_2 is considered. The cost estimate is limited to installed cost of listed equipment. It does not include e.g. land procurement, preparation, service buildings or owners cost.

2.4.2 Dimensioning of equipment

For the absorber and desorber internals, a structured packing was chosen because it yields a low pressure drop, high efficiency and high capacity (Øi, 2012). To determine the packing height, a constant stage (Murphree) efficiency corresponding to 1 meter of packing was assumed. Murphree efficiencies of 0.15 and 1.0 were specified for the absorber and the desorber in Table 1.

Centrifugal pumps with 75 % adiabatic efficiency were used in the process simulation. The pump duties that are calculated by Aspen HYSYS does not take the lifting height into consideration, only the pressure difference across the pump. Since the pumps are not the most expensive equipment in a CO_2 capture plant, an approximate additional duty compensating for the lifting height was included.

The absorption column diameter was calculated based on a gas velocity of 2.5 m/s and the desorption column is based on a gas velocity of 1 m/s as in Park and Øi (2017). The packing height of the absorption and desorption column is 1 meter per stage with a specified stage efficiency. The total height of the absorption column and desorption column is specified to be 40 m and 16 m respectively. The extra height is due to distributors, water wash packing, demister, gas inlet, outlet and sump.

Overall heat transfer coefficient values have been specified for lean/rich heat exchanger 550 W/(m²K), lean amine cooler 800 W/(m²K), reboiler 2500 W/(m²K) and condenser 2000 W/(m²K). These values are higher than in Øi (2012) and Park and Øi (2017) which are regarded as conservative.

2.4.3 Capital cost estimation methods

Table 2 specifies general assumptions made for the cost comparison.

Figure 2. Aspen HYSYS flow-sheet of the base case simulation

The equipment costs are taken from the Aspen In-plant Cost Estimator (v.10), which gives the cost in Euro (€) for Year 2016 (1st Quarter). A generic location that has good infrastructure and easy access to a workforce and materials, e.g. Rotterdam, is assumed. Stainless steel (SS316) with a material factor of 1.75 was assumed for all equipment units.

In the detailed factor method, each equipment cost (in carbon steel) was multiplied with its individual installation factor to get equipment installed cost, as in earlier works (Øi, 2012; Park and Øi, 2017). The total capital cost was then calculated by adding all the individual equipment installed costs. The detailed installation factor is a function of the site, equipment type, materials, size of equipment and includes direct costs for erection, instruments, civil, piping, electrical, insulation, steel and concrete, engineering cost, administration cost, commissioning and contingency. The updated installation factors for year 2016 (Eldrup, 2016) were used.

Table 2. Cost calculation specifications

Parameter	Value
Plant lifetime	20 years
Discount rate	7.5 %
Maintenance cost	5 % of installed cost
Electricity price	0.5 NOK/kWh
Steam price	0.13 NOK/kWh
Annual operational time	8000 hours
Location	Rotterdam
Currency exchange rate 2016	9.209 (European Central Bank, 2019)
Cost index 2016	103.6 (Statistics Norway, 2019)
Cost index September 2019	111.1

This cost estimate is expected to have an accuracy of ±40%.

2.4.4 Operating cost calculation

The sum of all the costs for running the project is calculated to be the total OPEX per year. This project includes OPEX estimations for the use of electricity and steam to run the CO_2 capture process. Electricity cost was specified to be 0.5 NOK/kWh (approximately 0.05 Euro/kWh). The steam cost was specified to be 25 % of the electricity cost, 0.125 NOK/kWh. Running uptime for the project was assumed to be 8000 hours.

2.4.5 Aspen HYSYS spreadsheet calculations.

The spreadsheet unit in Aspen HYSYS was used to calculate the detailed cost estimation of CAPEX, OPEX and NPV (net present value). The NPV was calculated as the sum of CAPEX and OPEX for a calculation period of 20 year, and with discount factor 7.5 % as specified in Table 2.

For the different alternatives (especially when varying the parameters to be optimized) the spreadsheet calculated the NPV for each set of alternative parameter values.

For the case of optimizing the temperature difference in the main heat exchanger, the calculation could be made effectively by using a Case Study option in Aspen HYSYS, so that the calculations could be performed automatically for each pre-selected parameter value.

For the case of optimizing the number of absorber stages, each calculation was performed independently by specifying the number of stages in each calculation. The optimum number of stages can then be found as the number giving the lowest NPV.

3 Results and Discussion

3.1 Simulations and convergence

The calculation sequence including the recycle block and the adjust operations were developed as a result of a combination of earlier work and trial and error. The recycle block did not converge without adjusting the makeup water (and makeup amine). This was done manually or it was calculated by a material balance. An improved procedure for the calculation of the water (and amine) makeup can probably increase the convergence efficiency.

The DeltaTmin adjust block was efficient to find the specified minimum temperature approach in the main heat exchanger. When the tolerance was reduced, more iterations (than the specified) was often necessary to achieve convergence. The Lean amine flow adjust block was the most difficult to converge. This can be explained by that the recycle block had to be solved for each iteration. The recycle block was difficult to converge by itself, and the variation of the amine flow to obtain the specified CO_2 removal grade made the convergence more difficult. When the tolerance was reduced, the number of iterations both in the recycle block and in the adjustment operation increased.

3.2 General optimization results

The results given in Figure 3, Figure 4 and Figure 5 are similar to results in an earlier report (Park et al. 2015) with similar conditions. The optimum parameter values are also close to values from earlier work (Kallevik, 2010; Øi, 2012; Aromada and Øi, 2017). Some test calculations indicate that the change in equilibrium model in Aspen HYSYS to the acid gas package, does not change the results much.

3.3 Optimum number of stages

Figure 3 shows the energy consumption (in MJ/kg CO_2 captured) and negative NPV (in mill. NOK) for the number of stages 8, 10 and 12.
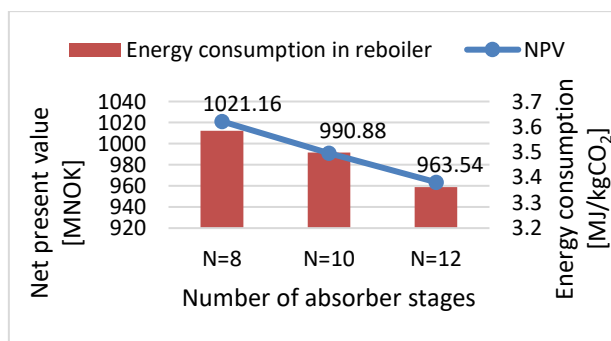
Figure 3. Energy consumption and negative net present value for the number of stages case simulations.

The figure shows that the energy consumption decreases significantly when the number of stages

increases. It also shows that the NPV becomes less negative as the number of stages increases. The decrease from 10 to 12 stages is only 3 %. This indicates that the optimum is slightly higher than 12. To optimize the number of stages, it is necessary to start a new calculation for every specified number of stages.

3.4 Optimum minimum T approach

Figure 4 shows the energy consumption (in MJ/kg CO_2 captured) and negative NPV (in mill. NOK) for the minimum temperature approach in the main amine heat exchanger equal to 5, 10, 13 and 15. The figure shows that the energy consumption increases significantly when the minimum temperature difference increases. It also shows that the NPV is very little dependent on the minimum temperature approach. When the conditions are changed slightly, the optimum temperature approach changes between approximately 10 - 15 ºC. This is similar to results from other work. To optimize the minimum temperature, the most efficient way found, is to perform a Case Study in Aspen HYSYS. Then the optimum value can be found as the one with minimum (negative) NPV. In principle, the optimization could be performed by adding a minimization procedure in the Aspen HYSYS spreadsheet, and return a next minimum temperature approach to the Aspen HYSYS program.
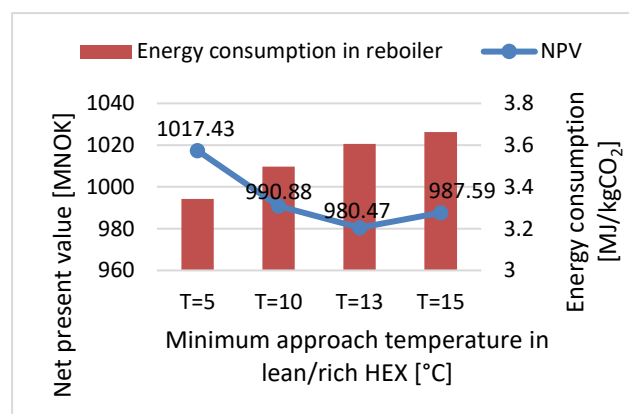
Figure 4. Energy consumption and negative net present value for the minimum approach temperature case.

3.5 85 or 90 % CO_2 removal

Figure 5 shows the specific capture cost (in NOK/ton CO_2 captured) for 85 % and 90 % removal. The cost is distributed on CAPEX, electricity and steam. The figure shows that the steam cost is dominating. 85 % removal gives the minimum specific capture cost. It is however a strategic choice whether 85 or 90 % should be selected because it is not obvious whether a high removal CO_2 or a low specific CO_2 capture cost should be chosen. Compared to more detailed cost estimates on CO_2 capture cost, the calculated cost is probably underestimated (Park et al., 2015). In principle, such an optimization could be performed using a minimization procedure in the Aspen HYSYS spreadsheet.
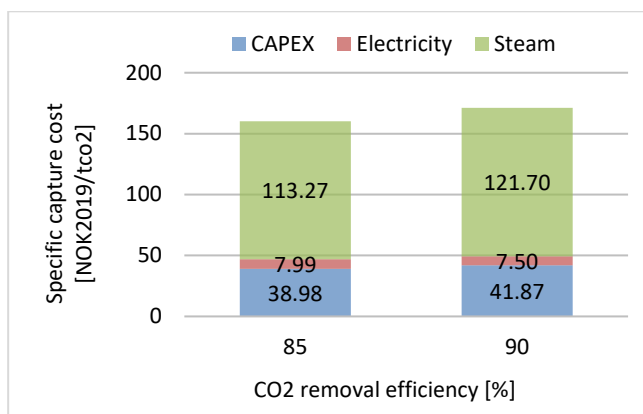
Figure 5. Specific capture cost [NOK/ton CO₂] for 85% and 90 % CO$_2$ capture.

4 Conclusion

Aspen HYSYS simulations of a standard amine based process for CO$_2$ capture using an equilibrium based model have been performed in Aspen HYSYS version 10.0 using flue gas data from a cement plant.

The capital cost of CO$_2$ capture was estimated using equipment cost data from Aspen In-plant and then using a detailed factor method. The cost analysis was limited to the absorption and circulation system, and CO$_2$ compression or liquefaction was not included. Operating cost was estimated from calculated electricity and heat consumption, and maintenance cost was based on estimated capital cost. Parameters varied were the minimum temperature difference in the main heat exchanger, the number of absorption stages and % CO$_2$ removed in the process.

Optimum temperature difference in the main heat exchanger was calculated to 10-15 °C, dependent on the specifications. This was found after one simulation for each temperature. Optimum column height was calculated with 12 stages (equivalent to 12 meter of structured packing) based on one simulation for each stage number. Compared to more detailed cost estimates on CO$_2$ capture cost, the calculated cost of 180-190 NOK (18-19 Euro) is probably underestimated. The scope of the cost calculation is limited to the absorption and circulation system which is most important for parameter optimization.

To obtain really automated calculations it is recommended to improve the robustness of the simulations. This may be achieved by making the material balances more accurate. It should in principle be possible to optimize e.g. the temperature difference in only one automated calculation. To optimize the height (number of stages) in the absorption column automatically, a way to update the number of stages during the simulations has to be found.

References

H. Ali. *Techno-economic analysis of CO₂ capture concepts.* PhD Thesis, University of South-Eastern Norway, 2019.

S. A. Aromada and L. E. Øi. Simulation of Improved Absorption Configurations for CO₂ Capture. In *Linköping Electronic Conference Proceedings SIMS 56,* pages 21-29, 2015. doi:http://dx.doi.org/10.3384/ecp1511921

S. A. Aromoda and L. E. Øi. Energy and Economic Analysis of Improved Absorption Configurations for CO₂ Capture. *Energy Procedia*, 114:1342-1351, 2017.

N. H. Eldrup. *Installation factor sheet - Project Management and Cost Engineering*. Master's Course. University College of Southeast Norway, Porsgrunn, 2016.

European Central Bank. *ECB euro referance exchange rate: Norwegian krone (NOK)*. Available 3.11.2019: https://www.ecb.europa.eu/stats/policy_and_exchange_rates/euro_reference_exchange_rates/html/eurofxref-graph-nok.en.html]

A. L. Haukås, J. Helvig, I. Hæstad, and A. M. Lande. *Automatization of Process Simulation and Cost Estimation of CO₂ capture in Aspen HYSYS*. Master's Project, University of South-Eastern Norway, Porsgrunn, 2019.

J. Holoboff. *Improving flowsheet convergence in HYSYS*. Available online (30.11.2019): http://processecology.com/articles/improving-flowsheet-convergence-in-hysys

Y. Ishii and F. D. Otto. Novented process plowsheeting Part I: A basic algorithm for inter-linked, multicolumn separation processes. *Computers and Chemical Engineering*, 32:1842-1860, 2008.

O. B. Kallevik. *Cost estimation of CO₂ removal in HYSYS*. Master's Thesis, Telemark University College, Porsgrunn, 2010.

T. P. Kisala, R. A. Trevino-Lozano, J.F. Boston, and H. I. Britt. Sequential modular and simultaneous modular strategies for process flowsheet optimization. *Computers and Chemical Engineering*, 11(6):567-579, 1987.

Norcem. *Sementproduksjon og CO2*. Available 16.9.19: https://www.norcem.no/no/sementproduksjon-co2

K. G. Park, S. Hartantyo, A. Nazari, T. Huang, and B. Rai. *Process simulation and cost estimation of CO₂ capture from cement plant using Aspen HYSYS*. Master's Project, Telemark University College, Porsgrunn, 2015.

K. Park and L. E. Øi. Optimization of gas velocity and pressure drop in CO₂ absorption column. In *Linköping Electronic Conference Proceedings SIMS 58,* pages 292-297, 2017. doi: 10.3384/ecp17138292

A. B. Rao and E. S. Rubin. Technical, Economic and Environmental Assessment of Amine-based CO₂ Capture Technology for Power Plant Greenhouse Gas Control. *Environmental Science and Technology*, 36:4467-4475, 2002. doi: 10.1021/es0158861

Statistics Norway. *Konsumprisindexen*. Available 3.11.2019: https://www.ssb.no/priser-og-prisindekser/statistikker/kpi

L. E. Øi. Aspen HYSYS simulation of CO₂ removal by amine absorption in a gas based power plant. In Proceedings *The 48th Scandinavian Conference on Simulation and Modelling (SIMS 2007)*, Göteborg, Sweden, 2007.

L. E. Øi. *Removal of CO₂ from exhaust gas*. PhD Thesis, Telemark University College, Porsgrunn. (TUC 3: 2012)

Simulation of CO₂ Absorption at TCM Mongstad for Performance Data Fitting and Prediction

Lars Erik Øi[1,*] Sofie Fagerheim[1,2]

[1]Department of and Process, Energy and Environmental Technology, University of South-Eastern Norway
[2]Multiconsult, N-8001 Bodø

lars.oi@usn.no

Abstract

The main purpose of this work has been to fit simulated models to performance data from Test Centre Mongstad (TCM), and evaluate whether fitted parameters for one scenario (a set of experimental data at specified conditions) give reasonable predictions at other conditions. Five scenarios from the amine based CO_2 absorption process at TCM have been simulated in a rate-based model in Aspen Plus and an equilibrium based model in Aspen HYSYS and Aspen Plus. It was evaluated whether a fitted interfacial area (for the rate-based model) or an E_M-profile (Murphree efficiency on each of 24 stages for the equilibrium based model) gave a good prediction of CO_2 removal rate and temperature profile for other conditions. An indication of the predictive performance of the rate-based model is that the interfacial area fitted to the different scenarios had to be varied between 0.29 and 1.0 to obtain the experimental CO_2 removal efficiency. Using a specific E_M-profile was able to predict both the CO_2 removal and the temperature profile for all the scenarios reasonably well. An E_M-factor multiplying all the E_M values in an E_M-profile from another scenario was fitted to values between 0.60 and 1.02 for all the scenarios.

Keywords: CO₂, amine, absorption, simulation

1 Introduction

The CO₂ Technology Centre Mongstad (TCM) close to Bergen is the world's largest test facility for CO₂ capture technology. For testing of CO₂ absorption into amine based solvents, there is an absorption column with a rectangular cross section equivalent to a packing diameter of 3 meter, and a packing height up to 24 meter. There have been performed performance tests of CO₂ absorption from flue gas into 30 wt-% monoethanol amine (MEA) in 2013 (Thimsen et al., 2014; Hamborg et al., 2014) and in 2015 (Gjernes et al., 2017; Faramarzi et al., 2017). Figure 1 shows the principle of the amine based CO₂ absorption and also the desorption facility at TCM. In this work, especially the total CO₂ capture rate (in % of incoming CO₂) in the absorption section and the temperature profile from top to bottom of the absorption section are the evaluated parameters.

The first aim of this work is to compare results from simulations with performance data for CO₂ absorption into 30 wt-% MEA at TCM using different simulation tools. The second aim, which is specific for this work, is to test whether fitted parameters for one scenario (a set of experimental data at specified conditions) give reasonable predictions at other conditions. The work is based on the Master Thesis of Sofie Fagerheim (2019) and on earlier work at the University of South-Eastern Norway.
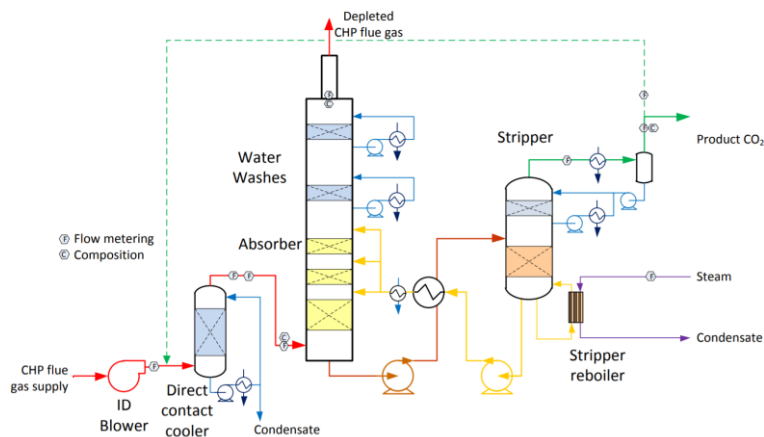
Figure 1. Simplified process diagram of the amine based CO₂ capture plant at TCM (Thimsen et al., 2014)

In earlier work (Sætre, 2016; Røsvik, 2018; Øi et al., 2018) the equilibrium models (in Aspen Plus and Aspen HYSYS) were fitted to one specific scenario by adjusting the Murphree efficiency (E_M) for each stage, and the rate-based model (in Aspen Plus) was fitted to another scenario by adjusting the interfacial area factor. In literature there is very little research on predicting CO_2 absorption with models fitted for other conditions.

In this work, 5 sets of experimental data (scenarios) from the amine based CO_2 capture process at TCM have been compared with simulations of different equilibrium based models and a rate-based model. The E_M-profile was then adjusted with an E_M-factor in the other scenarios to achieve a good fit to the temperature profile.

2 Process simulation tools and CO_2 absorption models

There are several process simulation tools available for CO_2 absorption processes. The key content in these programs are models for vapour/liquid equilibrium calculations and efficient flow-sheet solvers. Some of the programs (especially the programs with rate-based tools) also include models for chemical, heat transfer and mass transfer kinetics. Commercially available programs are Aspen Plus, Aspen HYSYS, ProTreat, ProMax and ChemCad. Some companies have internal programs, SINTEF use the program CO2SIM.

There are different equilibrium models used for the MEA/water/CO_2 system describing the relations between the vapour and liquid phase at equilibrium. Aspen Plus has an Electrolyte-NRTL equilibrium model which is based on Austgen et al. (1989). The new version of Aspen HYSYS has a new acid gas model. This work is mainly based on the earlier models in Aspen HYSYS which is based on the amine package with the Kent-Eisenberg (1976) and the Li-Mather (1994) equilibrium models.

Aspen Plus has included rate-based models. For CO_2 absorption, there are several models available for heat transfer, mass transfer and kinetics which can be included in a rate-based simulation. A specific rate-based example file for CO_2 removal using MEA is available with the Aspen Plus program. The parameters in this file are mostly based on the work of Zhang et al. (2009) who fitted Aspen Plus simulations to experimental runs at a CO_2 absorption pilot plant. Different rate-based models have been developed for TCM in the Master Thesis works of Larsen (2014), Desvignes (2015) and Sætre (2016). The gCCS program has been used for dynamic simulations at TCM (Bui et al., 2020).

Equilibrium based absorption models are based on the assumption of equilibrium at each stage. The model can be extended by using a Murphree efficiency (the ratio of the change in mole fraction from a stage to the next divided by the change assuming equilibrium). An advantage using Murphree efficiencies compared to rate-based simulations is that it is simpler and fewer parameters have to be specified. In the Master Thesis work of Zhu (2015), Sætre (2016) and Røsvik (2018), a Murphree efficiency for each stage (meter of packing) was estimated for one set (scenario) of TCM data (Hamborg et al., 2014). Zhu (2015) fitted a constant Murphree efficiency to 0.09 in Aspen HYSYS for all stages to obtain the experimental CO_2 capture rate. Using different fitted Murphree efficiencies for each stage, good agreement between the measured and simulated temperature profile was also obtained.

There are a few comparisons between different simulation tools for CO_2 absorption in literature. Luo et al. (2009) tested Aspen RadFrac, ProTreat, ProMax, Aspen RateSep, CHEMASIM from BASF and CO2SIM from SINTEF/NTNU by comparing with pilot plant data. The result was that all models were capable of fitting the CO_2 capture rate, but the temperature and concentration profiles were not well predicted. When comparing Aspen HYSYS and Aspen Plus, Øi (2012) claimed that there were small differences between the tested equilibrium models, and that a rate-based and equilibrium based model with estimated Murphree efficiencies gave similar results. In the work by Øi et al. (2018), different models were compared for 4 scenarios from TCM. The results showed that equilibrium and rate-based models perform equally well in both fitting performance data and in predicting performance at changed conditions.

3 Data, Methods and Specifications

3.1 Performance Data from TCM

Performance data for this work have been taken from 5 sets of conditions (scenarios) at TCM. They are from campaigns in 2013 and 2015 for approximately 30 wt-% MEA in water. 24 meter of packing height (the maximum available) was used in these scenarios.

The data (mainly conditions of the inlet gas stream and the inlet amine stream to the absorption section) for the 5 scenarios are listed in Table 1 to Table 5. Table 1 to 4 are for the same conditions as in Øi et al. (2018). The data are from scenarios documented in Hamborg et al. (2014), Gjernes et al. (2017) and Faramarzi et al. (2017), but some of the data are converted to make them suitable for input to simulation programs.

The 5 scenarios which have been selected in this work are named H14 and 6w from 2013 (Hamborg et al., 2014), 2B5 and Goal1 from 2015 (Gjernes et al., 2017) and F17 from 2015 (Faramarzi et al., 2017). The names have been used internally at TCM, except the H14 and F17 scenarios. The 5 scenarios were run with amine concentrations close to 30 wt-% MEA in water.

Table 1 Scenario H14 experimental input data for process simulations.

Input data to the simulations			
Amine inlet		**Flue gas inlet**	
Flow rate [kg/h]	54900	Flow [kmol/h]	2022
Temperature [°C]	36.5	Temperature [°C]	25.0
MEA [mol%]	10.94	CO_2 [mol%]	3.7
H_2O [mol%]	86.54	H_2O [mol%]	2.95
CO_2 [mol%]	2.52	O_2 [mol%]	13.6
Pressure [bara]	1.0313	N_2 [mol%]	79.75
		Pressure [bara]	1.063

Table 2 Scenario 6w experimental input data for process simulations.

Input data to the simulations			
Amine inlet		**Flue gas inlet**	
Flow rate [kg/h]	54915	Flow [kmol/h]	2005
Temperature [°C]	36.9	Temperature [°C]	25
MEA [mol%]	11.13	CO_2 [mol%]	3.57
H_2O [mol%]	86.37	H_2O [mol%]	3.0
CO_2 [mol%]	2.5	O_2 [mol%]	13.6
Pressure [bara]	1.0313	N_2 [mol%]	79.83
		Pressure [bara]	1.063

Table 3 Scenario 2B5 experimental input data for process simulations.

Input data to the simulations			
Amine inlet		**Flue gas inlet**	
Flow rate [kg/h]	49485	Flow [kmol/h]	2022
Temperature [°C]	36.8	Temperature [°C]	28.2
MEA [mol%]	11.67	CO_2 [mol%]	3.57
H_2O [mol%]	85.65	H_2O [mol%]	3.7
CO_2 [mol%]	2.68	O_2 [mol%]	14.6
Pressure [bara]	1.0313	N_2 [mol%]	78.08
		Pressure [bara]	1.063

Table 4 Scenario Goal1 experimental input data for process simulations.

Input data to the simulations			
Amine inlet		**Flue gas inlet**	
Flow rate [kg/h]	44391	Flow [kmol/h]	2017
Temperature [°C]	28.6	Temperature [°C]	25
MEA [mol%]	12.04	CO_2 [mol%]	3.62
H_2O [mol%]	85.19	H_2O [mol%]	3.1
CO_2 [mol%]	2.77	O_2 [mol%]	14.3
Pressure [bara]	1.0313	N_2 [mol%]	79
		Pressure [bara]	1.063

Table 5 Scenario F17 experimental input data for process simulations.

Input data to the simulations			
Amine inlet		**Flue gas inlet**	
Flow rate [kg/h]	57434	Flow [kmol/h]	2558
Temperature [°C]	37.0	Temperature [°C]	29.8
MEA [mol%]	11.44	CO_2 [mol%]	3.70
H_2O [mol%]	86.27	H_2O [mol%]	3.70
CO_2 [mol%]	2.29	O_2 [mol%]	14.60
Pressure [bara]	1.0313	N_2 [mol%]	78.00
		Pressure [bara]	1.0100

The results from the performance data scenarios to be compared with simulations, are the total CO_2 capture rate and the temperature profile measured for every meter of the packing. The temperature was measured at different locations for each meter of packing, and the different locations were named A, B, C and D.

For the H14 and F17 scenario, a mean value for each meter of packing was specified in the temperature profile (Hamborg, 2014; Faramarzi 2017).

3.2 Specifications for the Equilibrium Based Simulation Tools

When using Aspen HYSYS, the Amine Package with the Kent-Eisenberg model was used with non-ideal vapor phase. The Acid Gas model, which is the recommended equilibrium model in the last Aspen HYSYS version was used as a check. When using Aspen Plus, the Electrolyte-NRTL (Non-Random-Two-Liquid) model was used. The figures in this work show only results from Aspen Plus.

In the Master Thesis work of Zhu (2015), Sætre (2016) and Røsvik (2018) a Murphree efficiency for each of the 24 stages (meter of packing) was estimated for the TCM data set (Hamborg et al., 2014). Different approaches for fitting the E_M profile to the temperature profile was a constant E_M for every stage (Zhu, 2015) and a linear Murphree efficiency profile (Øi, 2012). A high Murphree efficiency at the top stages and then gradually reduced to a constant equal to 0.01 (close to 0) for the bottom stages fitted the temperature profile very well for the H14 scenario (Zhu, 2015; Sætre, 2017; Røsvik, 2018). Fagerheim (2019) fitted two other E_M profiles, SF1 and SF2 to the performance data. The SF1 and SF2 profiles has $E_M = 0.01$ on the 5 lowest stages compared to the Zhu profile which has $E_M = 0.01$ on the 10 lowest stages. These profiles were specified in both the Aspen HYSYS and Aspen Plus simulation tool. The different E_M profiles used in this work are presented in Table 6.

Table 6 Murphree efficienciy profiles used in this work

Murphree efficienciy profiles in the packed column from top to bottom					
EM	**0,1**	**Zhu**	**Lin**	**SF1**	**SF2**
1	0.1	0.2300	0.17	0,2450	0,2400
2	0.1	0.2192	0.17	0,2425	0,2350
3	0.1	0.2085	0.17	0,2400	0,2300
4	0.1	0.1977	0.17	0,2375	0,2250
5	0.1	0.1869	0.17	0,2350	0,2200
6	0.1	0.1800	0.16	0,2325	0,2150
7	0.1	0.1762	0.15	0,2300	0,2300
8	0.1	0.1546	0.14	0,2000	0,2000
9	0.1	0.1438	0.13	0,1700	0,1700
10	0.1	0.1331	0.12	0,1400	0,1400
11	0.1	0.1223	0.11	0,1100	0,1100
12	0.1	0.1115	0.10	0,0800	0,0800
13	0.1	0.1007	0.09	0,0500	0,0550
14	0.1	0.0900	0.08	0,0475	0,0525
15	0.1	0.0100	0.07	0,0450	0,0500
16	0.1	0.0100	0.06	0,0425	0,0475
17	0.1	0.0100	0.05	0,0400	0,0450
18	0.1	0.0100	0.04	0,0375	0,0425
19	0.1	0.0100	0.03	0,0350	0,0400
20	0.1	0.0100	0.02	0,0001	0,0001
21	0.1	0.0100	0.01	0,0001	0,0001
22	0.1	0.0100	0.01	0,0001	0,0001
23	0.1	0.0100	0.01	0,0001	0,0001
24	0.1	0.0100	0.01	0,0001	0,0001

3.3 Specifications for the Rate-Based Tool

The specifications in the rate-based Aspen Plus simulation tool at TCM have been developed during several years and different versions have been used (Larsen, 2014; Desvignes, 2015; Sætre 2016). In this work, the same parameters as in Øi et al. (2018) were used. Most of these specifications are based on the work by Zhang et al. (2009) where Aspen Plus rate-based simulations were fitted to pilot scale experiments of CO_2 absorption at the University of Texas. Detailed documentation of the rate-based model can be found in the Aspen Plus program documentation.

4 Results and discussion

4.1 General Results

The results shown for each model in the scenario figures are the capture rate and the temperature profile. The model parameters (in the case of the rate-based model the interfacial area factor) are adjusted to achieve the specified capture rate. In the case of using an E_M profile, all the E_M values were multiplied with an E_M-factor which is the only parameter. In some cases it was not possible to obtain the specified capture rate. In that case the parameter was adjusted to come as close as possible to the specified capture rate. The emphasis in this work is on comparison of the temperature profiles.

4.2 Scenario H14

Comparisons between measured and simulated temperature profiles are shown in Figure 2. In the figure, the number after each model is the E_M-factor (adjusted to achieve the capture rate given in the parenthesis). The broad solid line is representing the experimental data as an average of 4 measured temperatures at each stage. The non-smooth form of the experimental line indicates that there is some uncertainty in the measurements.
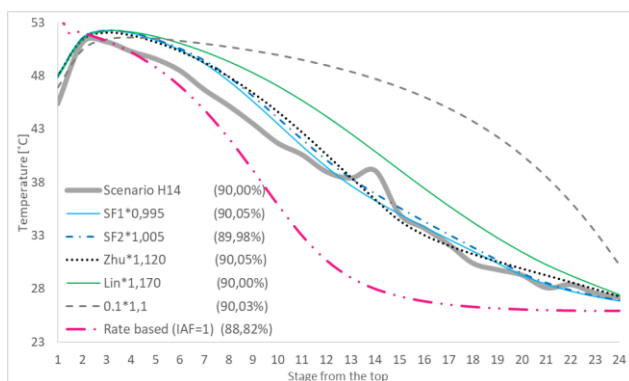
Figure 2. Comparison of plant data scenario H14 and simulated temperature profiles (Fagerheim, 2019).

The model with constant E_M overpredicts the temperature to a large degree in the lower part of the column. The rate-based model (with interfacial area 1.0) underpredicts the temperature. The model with a linear E_M profile gives a reasonable temperature profile. The adjusted E_M profiles give the best fit. This is as expected because all the E_M values (24 parameters) are actually fitted to this scenario. In the work of Øi et al. (2018), the rate-based model gave a temperature profile closer to the experimental line for the H14 scenario using Aspen HYSYS. In that case, an interfacial area of 0.55 was used, but this does not achieve the capture rate in the Aspen Plus simulations. Even with an interfacial area factor of 1.0, the Aspen Plus simulations only achieved a capture rate slightly below the experimental value of 90 %. The slight difference between the results using Aspen HYSYS and Aspen Plus, is probably due to different vapour/liquid equilibrium models.

4.3 Scenario 6w

Comparisons between measured and simulated temperature profiles are shown in Figure 3. The experimental temperatures are shown in the figure, and the broad solid line shows the average. It is clear that the uncertainty in the experimental temperatures are not negligible.

Figure 3. Comparison of plant data scenario 6w and simulated temperature profiles (Fagerheim, 2019).

In this scenario all the models except the constant E_M model give reasonable temperature profiles. In this case, the rate-based model overpredicts the temperature in the lower part of the column. The interfacial area had to be adjusted to 0.29 to achieve the experimental capture rate. The models with E_M profiles fitted to the H14 scenario, had to be adjusted with factors from 0.60 to 0.72 to obtain the experimental capture rate.

4.4 Scenario 2B5

Comparisons between measured and simulated temperature profiles are shown in Figure 4. The figure shows a higher uncertainty in the experimental temperatures than the other scenarios.
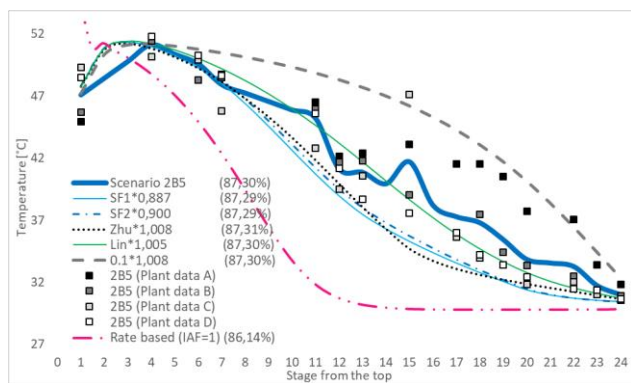
Figure 4. Comparison of plant data scenario 2B5 and simulated temperature profiles (Fagerheim, 2019).

As in the H14 scenario, the model with constant E_M overpredicts the temperature in the lower part of the column and the rate-based model underpredicts the temperature. The interfacial area in the rate-based model had to be adjusted to 1.0 to (almost) achieve the experimental capture rate.

The E_M profile models give a good fit with an adjustment factor between 0.89 and 1.01. The uncertainty in the data is too large to conclude which model is closest to the experimental data.

4.5 Scenario Goal1

Comparisons between measured and simulated temperature profiles are shown in Figure 5. In this scenario all the models overpredicts the temperature in the middle of the column. The rate-based model and the linear model overpredicts the temperature with about 10 K. The E_M profile models overpredicts the temperature with less than about 3 K. The interfacial area was adjusted to 0.51, while the other models were adjusted with factors between 0.90 and 1.02.
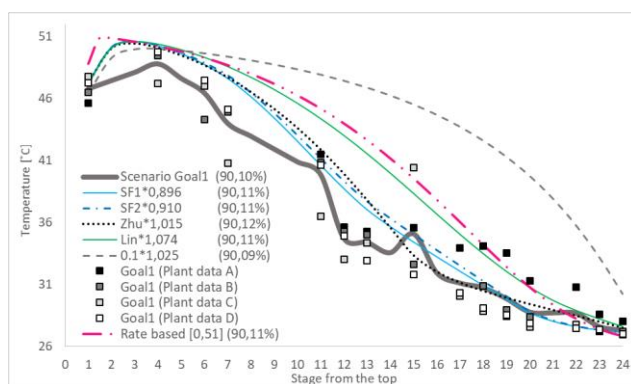
Figure 5. Comparison of plant data scenario Goal1 and simulated temperature profiles (Fagerheim, 2019).

4.6 Scenario F17

Comparisons between measured and simulated temperature profiles are shown in Figure 6. The broad solid line is representing the experimental data. The experimental line which is smoother than in the earlier scenarios indicates that the experimental uncertainty is lower in the F17 scenario. There is however an outlier in the experimental temperature on stage 14 in all the scenarios. The linear model is closest to the experimental data for this scenario. The rate-based model overpredicts the temperature slightly in the middle part of the column.
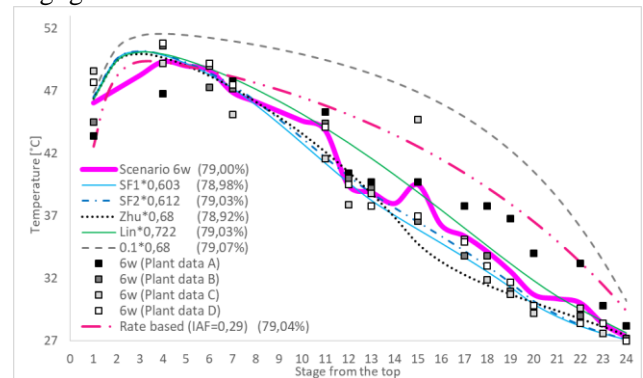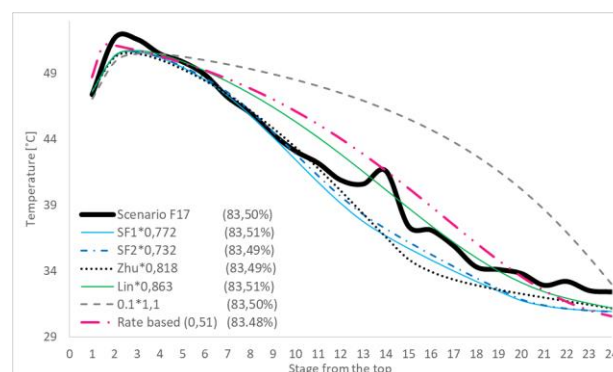

Figure 6. Comparison of plant data scenario F17 and simulated temperature profiles (Fagerheim, 2019).

The E_M models underpredicts the temperature with up to 3 K in the lower part of the column. The interfacial area was adjusted to 0.51, while the other models were adjusted with factors between 0.72 and 0.86.

4.7 General discussion

Only the results using the Electrolyte-NRTL model in Aspen Plus were shown in the figures in this work. Earlier work has shown that different equilibrium models give similar results (Sætre, 2016; Øi et al., 2018), and this has also been confirmed in simulations in Fagerheim (2019).

Some references (Zhang et al., 2009; Larsen, 2014; Desvignes, 2015) have compared also the concentration profiles when comparing performance data and simulation tools. This may give additional information for analysis.

Only performance data with 24 meter of packing was used in this and earlier work. The low Murphree efficiency on the lowest stages indicate that the number of stages in the column is too high. Equilibrium based and rate-based models could also be fitted to performance data for lower packing heights which are available for e.g. 18 meter packing height (Gjernes, 2017).

This work indicates that equilibrium based models give better predictions at other conditions. This is probably not a general conclusion. Equilibrium based models are very empirical and are probably only useful when the conditions do not change much. Some have claimed (Zhang et al., 2009) that rate-based models are superior to equilibrium based models. There are several factors and parameters in the rate-based models which are not well known, typically specifications for fluid flow, heat transfer and mass transfer mechanisms in structured packings. When the knowledge of these factors becomes better known, the rate-based models can probably be made more predictive.

5   Conclusions

Five different scenarios from the CO_2 capture process at TCM have been simulated in a rate-based model in Aspen Plus and in an equilibrium-based model in Aspen HYSYS and Aspen Plus. In the rate-based model, the performance data was fitted by changing only the interfacial area factor to obtain the experimental CO_2 removal efficiency. The simulated temperature profile from top to bottom of the absorption column was then only qualitatively correct compared to the measured temperature profile. In the equilibrium based model, a Murphree efficiency (E_M) was specified for each of 24 stages (meter of packing) to fit both the CO_2 removal efficiency and the temperature profile for one scenario. In this work different E_M-profiles (different E_M values on each stage) were examined to fit the temperature profile for a given scenario. The E_M-profiles were then used to fit performance data for other scenarios by adjusting only an E_M-factor which multiplies all the E_M values in an E_M-profile.

It was evaluated whether a fitted interfacial area (for the rate-based model) or a fitted E_M-profile (for the equilibrium based model) for one scenario gave a good prediction for other scenarios. The rate-based model fitted for a certain scenario was not able to predict performance well for all other scenarios. An indication of this was that the interfacial area in the different scenarios had to be varied between 0.29 and 1.0 to obtain the measured CO_2 removal efficiency. Using a E_M-profile fitted for a specific scenario was able to predict performance better for all the scenarios. By multiplying the specified E_M-profile with an E_M-factor (only one parameter), the fit at a new scenario was reasonably accurate. The fitted E_M-factor for each scenario had to be varied between 0.60 and 1.02 to obtain the measured CO_2 removal for all the scenarios.

The performance (CO_2 removal efficiency and temperature profile) was reasonably fitted and simulated for each scenario by all the models. None of the models are however expected to predict accurate performance for conditions far from the fitted scenario without any additional adjustment.

References

D. M. Austgen, G. T. Rochelle, X. Peng and C. Chen. Model of Vapor-Liquid Equilibria for Aqueous Acid Gas-Alkanolamine Systems Using the Electrolyte-NRTL Equation. *Industrial & Engineering Chemistry Research,* 28:1060-73, 1989.

M. Bui, N. E. Flø T. de Cazanove, and N. Mac Dowell. Demonstrating flexible operation of the Technology Centre Mongstad (TCM) CO_2 capture plant. *International Journal of Greenhouse Gas Control,* 93:1-26, 2020.

C. Desvignes. *Simulation of post-combustion CO_2 capture process with amines at CO_2 Technology Centre Mongstad.* Master's Thesis, CPE Lyon, France, 2015.

S. Fagerheim. *Process simulation of CO_2 absorption at TCM Mongstad.* Master's Thesis. University of South-Eastern Norway, 2019.

L. Faramarzi, D. Thimsen, S. Hume, A. Maxon, G. Wattson, S. Pedersen, E. Gjernes, B. F. Fostås, G. Lombardo, T. Cents, A. K. Morken, M. I. Shah, T. de Cazenove and E. S. Hamborg. Results from MEA testing at the CO_2 Technology Centre Mongstad: Verification of baseline results in 2015. *Energy Procedia,* 114:1128-1145, 2017.

E. Gjernes, S. Pedersen, T. Cents, G. Watson, B. F. Fostås, M. I. Shah, G. Lombardo, C. Desvignes, N. E. Flø, A. K. Morken, T. de Cazenove, L. Faramarzi and E. S. Hamborg. Results from 30 wt% MEA performance testing at the CO_2 Technology Centre Mongstad. *Energy Procedia,* 114:1146-1157, 2017.

E. S. Hamborg, V. Smith, T. Cents, N. Brigman, O. Falk-Pedersen, T. de Cazenove, M. Chagnlal, J. K. Feste, Ø. Ullestad, H. Ulvatn, O. Gorset, I. Askestad, L. K. Gram, B. F. Fostås, M. I. Shah, A. Maxson and D. Thimsen. Results from MEA testing at the CO_2 Technology Centre Mongstad. Part II: Verification of baseline results. *Energy Procedia,* 63:5994-6011, 2014.

R. L. Kent, B. Eisenberg. Better data for Amine Treating. *Hydrocarbon Processing,* 55(2):87-90, 1976.

I. M. S. Larsen. *Simulation and validation of CO_2 mass transfer processes in aqueous MEA solutions with Aspen plus at TCM.* Master's Thesis, Telemark University College, Norway, 2014.

Y. Li, A. E. Mather. Correlation and Predicition of the Solubility of Carbon Dioxide in a Mixed Alkanol Solution. *Industrial & Engineering Chemistry Research,* 33:2006-2015, 1994.

X. Luo, J. N. Knudsen, D. de Montigny, T. Sanpasertparnich, R. Idem, D. Gelowitz et al. Comparison and validation of simulation codes against sixteen sets of data from four different pilot plants. *Energy Procedia,* 1:1249-56, 2009.

O. Røsvik. *Process simulation of CO_2 capture at Mongstad.* Master's Thesis. University of South-Eastern Norway, 2018.

K. A. Sætre. *Evaluation of process simulation tools at TCM.* Master's Thesis. University College of South-Eastern Norway, 2016.

D. Thimsen, A. Maxson, V. Smith, T. Cents, O. Falk-Pedersen, O. Gorset and E. S. Hamborg. Results from MEA testing at the CO_2 Technology Centre Mongstad. Part I: Post-Combustion CO2 capture testing methodology. *Energy Procedia,* 63:5938-5958, 2014.

Y. Zhang, H. Chen, C. Chen, J. Plaza, R. Dugas, G. T. Rochelle. Rate-based Process Modelling Study of CO_2 Capture with Aqueous Monoethanolamine Solution. *Industrial & Engineering Chemistry Research,* 48:9233-46, 2009.

Y. Zhu. *Simulation of CO_2 capture at Mongstad using Aspen HYSYS.* Master's Thesis. Telemark University College, Norway, 2015.

L. E. Øi. Comparison of Aspen HYSYS and Aspen Plus simulation of CO_2 absorption into MEA from atmospheric gas. *Energy Procedia,* 23:360-369, 2012.

L. E. Øi, K. A. Sætre and E. Hamborg. Comparison of Simulation Tools to Fit and Predict Performance data of CO_2 Absorption into Monoethanol Amine. In *Linköping Electronic Conference Proceedings SIMS 59,* pages 230-35, 2018. doi: http://dx.doi.org/10.3384/ecp18153230

The Effect of Input Distribution Skewness on the Output Distribution for Total Project Schedule Simulation

Krige Visser

Department of Engineering and Technology Management, University of Pretoria, South Africa
`krige.visser@up.ac.za`

Abstract

A Monte Carlo simulation is useful to determine the probability of completing a project within budget and on time at various stages of the project. This paper discusses a research study to determine and compare the distribution for total project duration using ten different probability distributions for fourteen activities of a project. Triangular distributions were used as reference and parameter values for each activity duration were assumed. Parameter values of nine other distributions were calculated from the mean and standard deviations for the triangular distribution. The study indicated that P90 and P95 values of the output distributions, using different input distributions, differed by up to 4,8%. It was also found that there is a positive correlation between the mean skewness of the fourteen activities and the P90 and P95 values. The skewness of the output distribution showed a strong positive correlation with the mean skewness of the input distributions.

Keywords: risk, simulation, duration, projects

1 Introduction

1.1 Background

Monte Carlo simulation is mentioned in the Project Management Body of Knowledge (PMBoK) guide (PMI, 2014) as one of the quantitative risk analysis tools. A detailed discussion of cost and schedule risk analysis and simulation is provided by Cooper et al. (2014). A schedule risk simulation is typically done before project implementation and can be updated with actual activity duration values to replace the input distributions as the project is executed. The probability of finishing the project by the due date can thus be captured at various stages or phases of the project and plotted on a timeline to determine the trend.

Two notable projects in which schedule simulation was used extensively are the Øresund bridge project in Denmark and Sweden (Christensen and Rydberg, 2001), and the Gotthard base tunnel in Switzerland (Ehrbar et al., 2016). In the former the bridge and tunnel was completed 5 months ahead of schedule and in the latter the 57km tunnel underneath the Alps was completed a year earlier than initially planned.

It is unusual for a person to perform the same activity in exactly the same time when repeated, therefore, probability distributions are used to describe this uncertainty in duration. Various distributions have been proposed to model the uncertainty in the duration of project activities, e.g. the betapert, triangular, normal and lognormal distributions (Raydugin, 2013; Munier, 2014). The normal distribution is symmetric, the lognormal distribution is right skewed and the betapert and triangular distributions can be left skewed, symmetric, or right skewed.

If all the activity durations of a project are modelled with the normal distribution, one would expect the total project duration to be symmetric as well. This is the case if all activities are performed in series. If some activities are performed in parallel, the output distribution could be skewed even if all input distributions are symmetric. If all the activities are modelled with a right skewed distribution one would expect the distribution for the total project duration to also be right skewed. The skewness of the input distributions would therefore determine the skewness of the output distribution. Schedule simulation is mostly used to determine the 90% (P90) or 95% (P95) certainty duration of the project and the skewness of the output distribution affects the value of the P90 and P95 values. The choice of input distribution is therefore important to obtain a good approximation of the P90 and P95 values.

1.2 Objectives of Study

The main objective of this study was to determine the effect of the skewness of input distributions of a project network on the skewness of the output distribution provided by a Monte Carlo simulation in Excel. A further objective was to investigate the differences in the skewness of the output distributions for ten different input distributions that have different skewness values for fourteen project activities. A project network with 14 activities in series and parallel was selected as a case study for this research project.

2 Literature

2.1 Schedule Risk

Several high-profile projects have suffered delays and eventual slippage. A notable example of severe slippage is the construction of the Sydney Opera House that took 10 years longer to finish (Steyn et al., 2016). Another classical example is the Central Artery/Tunnel project in Boston, USA. The project duration was planned for about 10 years but eventually took 17 years to complete (National Research Council, 2003).

Managing the schedule for a project is of great importance, especially for mega projects and very long projects. Slippage in a large or long transport project will inevitably lead to cost overruns and loss of revenue since toll fees cannot be collected (Love et al. 2014).

Vanhoucke (2015) mentioned the importance of schedule risk analysis in projects and said a sound baseline schedule is critical for successful execution of the subsequent phases of the project. Nicholas and Steyn (2017) said "project scheduling is an integral part of project planning".

Uncertainty in the total duration of a project is influenced by two factors. The first is uncertainty in the duration of individual activities and tasks to be performed. This type of uncertainty is treated by means of suitable probability distributions. The second uncertainty is due to the occurrence of random events that impact the duration of an activity. This type of uncertainty can be incorporated in a simulation by means of a probability of occurrence and estimates for the consequence should the event occur. Discrete distributions like the binomial distribution is useful for incorporating random events in a schedule simulation (Damnjanovic and Reinschmidt, 2020).

2.2 Probability Distributions

Various probability distributions are available to model activity duration in projects as well as operational and maintenance tasks in an enterprise. Only distributions that have two parameters were selected for this study, except the triangular distribution which has three parameters. Some information on four popular probability distributions are given below.

The triangular distribution is a versatile continuous distribution that is appropriate for cost and schedule simulations (Scherer et al., 2003). The parameters are easy to estimate but one drawback is the fact that the maximum duration of an activity is capped. Even if very skewed values are selected for the triangular distribution, the inverse of the cumulative distribution function cannot produce a higher value than the maximum or upper bound value.

The lognormal distribution is quite popular for cost and schedule risk simulation and was used extensively by Bowden et al. (2001). It is a right-skewed distribution that is often used to model the duration of activities that

are performed by novice artisans or technicians that perform complex tasks.

The normal distribution is probably the most popular continuous distribution used in risk simulation. It is symmetric and the parameters are quite easy to estimate, especially if similar activities have been performed in the project organization in the past.

The Fréchet distribution, also known as the inverse Weibull distribution, is an interesting distribution that is regarded as an extreme value distribution that is often applied to modeling extreme events. It is a "fat-tailed" distribution that can be useful to describe the uncertainty in the duration of new activities that have not been done before in a project organization.

Data on most of the probability distributions used in this study, i.e. formulae for the mean value, standard deviation, skewness, excess kurtosis, and the inverse variant, are provided by Evans et al. (2000).

2.3 Activity Duration

Many project network modelers use symmetric distributions to express uncertainty in activity duration, e.g. the normal distribution or logistic distribution. However, the distribution of activity durations in projects are often not symmetric but rather skewed to the right (positive skewness). According to Martens (2017), most activity distributions in projects are right skewed. This view is shared by Shankar (2011).

With many different distributions available to describe the uncertainty of activity duration, the question could be posed whether certain distributions are more suited for certain project situations. Various factors could influence this decision, e.g. ease of use, availability of inverse variants within add-in software, ease of estimating the parameters of the distribution, symmetry or not, positive inverse variants, and availability of an explicit formula to calculate the inverse variant if not available in add-in software.

Ferson et al. (1998) said "the results of probabilistic risk analyses are known to be sensitive to the choice of distributions used as inputs, an effect which is undoubtedly even stronger for the tail probabilities". The authors argue that, in the absence of information regarding the uncertainty of an input value, the uniform distribution with minimum and maximum bounds is the best approach.

Hajdu and Bokoro (2014) compared the results of simulations using the uniform, triangular and beta distributions as input. They found that the difference in the project duration for the three distributions was less than the effect on the duration due to a 10% variation in the values of the three-point estimates for the triangular distribution.

Sherer et al. (2003) approximated the normal distribution with a symmetric triangular distribution and approximated the lognormal distribution with a non-symmetric triangular distribution. The authors found

that the symmetric triangular distribution provides a good approximation for the normal distribution in the range of the mean $\pm 2{,}44\sigma$.

Wood (2002) compared the output cumulative distributions of a 12-activity project with two paths using the triangular, uniform, normal and lognormal distributions as input. He found that P90 values using different input distributions varied by as much as 10%.

Visser (2016) performed simulations for two project networks and compared the results for the triangular, normal, lognormal and betapert distributions. The P80 and P90 values of the output distribution with 20000 trials were compared. The study found no significant difference in the output for these four input distributions.

3 Methodology

A theoretical project network with 14 activities in series and parallel was chosen to study the effect of skewness of input distributions on the skewness of the total project duration. Project managers seem to agree that project activity durations mostly have right skewed distributions (Damnjanovic and Reinschmidt, 2020). The parameters of the triangular distributions for all activities were therefore chosen to provide positive skewness, i.e. right skewed distributions.

The mean and standard deviation values for the input distributions were used to determine the parameters of nine other distributions. Only distributions with two parameters were considered for this study. The betapert distribution that is popular in schedule simulation was excluded.

A model that considers the logic of the project network was developed in Excel. The schedule simulation was performed with 100000 trials using the SimVoi add-in for Excel (Treeplan, 2020). One output of the simulation was a chart of the cumulative distribution for the total project duration. The duration values at increments of 5% probability were also provided by the add-in. The project network that was used for this study is shown in Figure 1.
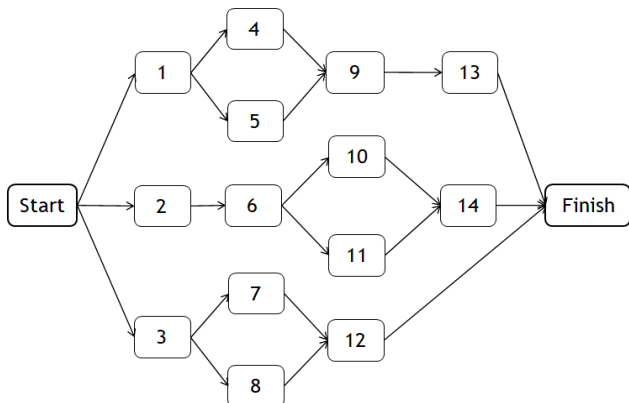
Figure 1. Project network used for simulation study

In this network there are six 'paths' that need to be executed in parallel to finish the total project. Values for the three parameters of the triangular distribution, i.e. a (lower bound), m (most likely) and b (upper bound), were chosen such that any of the six paths could be the critical path for the project. The mean value for each path of the network is shown in Table 1.

Table 1. Mean values of six paths in project network

Path	Lower bound	Most likely	Upper bound	Mean
1-4-9-13	27	35	62	41,33
1-5-9-13	25	32	57	38,00
2-6-10-14	26	33	58	39,00
2-6-11-14	28	35	59	40,67
3-7-12	28	35	54	39,00
3-8-12	29	35	53	39,00

The three parameter values for the triangular distribution, mean value (μ), standard deviation (σ) and the skewness (Skew) are shown in Table 2. The units for these parameters would typically be days or weeks if it is a long project, but it could represent any other time measurement.

Table 2. Input values for 14 activities

Act.	a	m	b	μ	σ	Skew
1	4	6	12	7,333	1,699	0,422
2	5	7	14	8,666	1,929	0,454
3	9	11	16	12,000	1,472	0,376
4	7	9	16	10,666	1,929	0,454
5	5	6	11	7,333	1,312	0,505
6	7	8	14	9,666	1,546	0,522
7	9	12	18	13,000	1,871	0,305
8	10	12	17	13,000	1,472	0,376
9	9	12	18	13,000	1,871	0,305
10	6	8	14	9,333	1,700	0,422
11	8	10	15	11,000	1,472	0,376
12	10	12	20	14,000	2,160	0,476
13	7	8	16	10,333	2,014	0,540
14	8	10	16	11,333	1,700	0,422
Mean Values				**10,762**	**1,725**	**0,426**

The mean and standard deviation values for each of the activities were used to calculate the parameters of the other probability distributions. Formulae for the mean and standard deviation of these nine distributions were mostly obtained from Evans et al. (2000) and the NIST e-Handbook of Statistical Methods (2012).

4 Results

4.1 Descriptive Statistics

Using the SimVoi add-in for Excel, the simulation provided the following values in Table 3 for the project duration distribution. The add-in provides inverse functions for most of the distributions and formulas were used for those not available. The truncated inverse function for the normal distribution was used to prevent negative duration values.

Table 3. Descriptive statistics for output distribution

Distribution	μ	σ	Skew	P90	P95
Fisk	43,719	3,03	+0,90	47,61	49,17
Frechet	43,771	3,49	+1,30	48,22	50,17
Gamma	43,723	2,76	+0,45	47,35	48,58
Gumbel	43,774	3,11	+0,74	47,87	49,43
Logistic	43,671	2,68	+0,49	47,15	48,38
Lognormal	43,742	2,84	+0,51	47,46	48,80
Normal	43,694	2,62	+0,33	47,13	48,22
Pareto	43,737	3,88	+1,77	48,59	50,89
Triangular	43,760	2,77	+0,32	47,41	48,56
Weibull	43,625	2,44	+0,20	46,80	47,80

The probability distributions in Table 3 are arranged alphabetically. The add-in provides the values for the mean value of the distribution (μ), the standard deviation (σ), skewness (Skew), P90, and P95. Other percentiles in increments of 5% are also provided. All the distributions produced a positive skewness for the output distribution. Even though all activities had a negative skewness for the Weibull distribution, the output distribution had a positive skewness.

4.2 Skewness of Total Project Duration

The skewness of the total project duration is illustrated graphically in Figure 2.
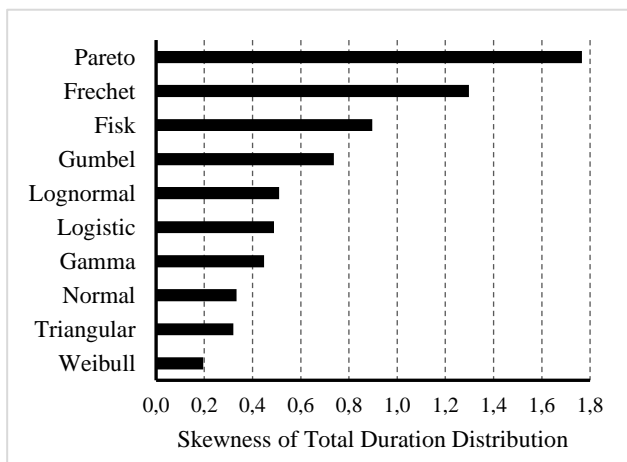
Figure 2. P90 and P95 values for ten input distributions

4.3 P90 and P95 Values

In some situations, a project manager might want to estimate the total duration with a 90% or 95% certainty (often when the project is nearing completion). The P90 and P95 values for the project duration using different input distributions are shown in Figure 3.
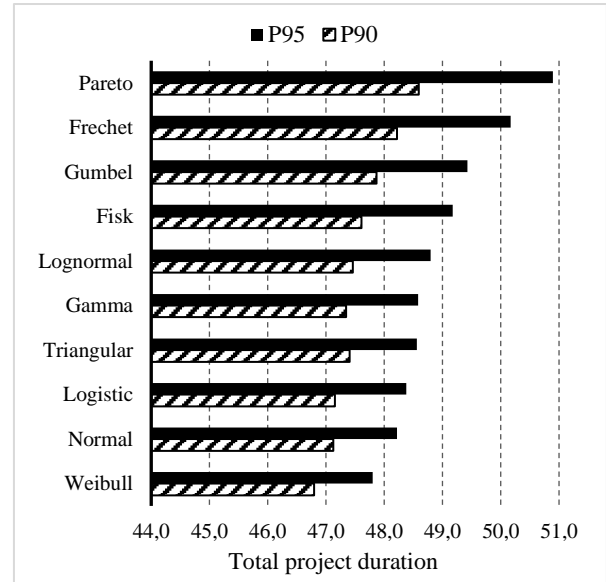
Figure 3. P90 and P95 values for 10 input distributions

The P95 values differed from 47,80 for the Weibull to 50,89 for the Pareto distribution. The value for the Weibull is 1,57% less than the value obtained for the triangular distribution while the value for the Pareto is 4,80% higher than that of the triangular. The Pareto is somewhat unpractical to describe the uncertainty in duration but the Fréchet distribution is well-known as a 'fat-tailed' distribution and is used in many engineering and scientific applications. The P95 value for the Fréchet distribution is 3,3% higher than the value of the triangular distribution. The P90 values for symmetric and slightly skewed distributions, i.e. the lognormal, gamma, logistic and normal distributions, did not differ significantly.

4.4 Effect of Skewness on P90 and P95

The effect of skewness of the input distributions of the activities was investigated. The mean skewness of all the activities was determined for all the distributions investigated and the relationship with the P90 and P95 values is shown in Figure 4.

It is evident from Figure 4 that there is a strong positive correlation between the skewness of the input distributions and the project duration as indicated by the P90 and P95 values. A correlation value of 0,94 was obtained for the relationship between the mean skewness of the input distributions and P90. A correlation value of 0,99 was obtained for the relationship between the mean skewness of the input distributions and P95. A least squares curve fit could provide an empirical formula for this relationship. This

will enable one to predict what the P90 and P95 values for the project duration will be for some other distribution if the mean skewness for that distribution can be calculated.
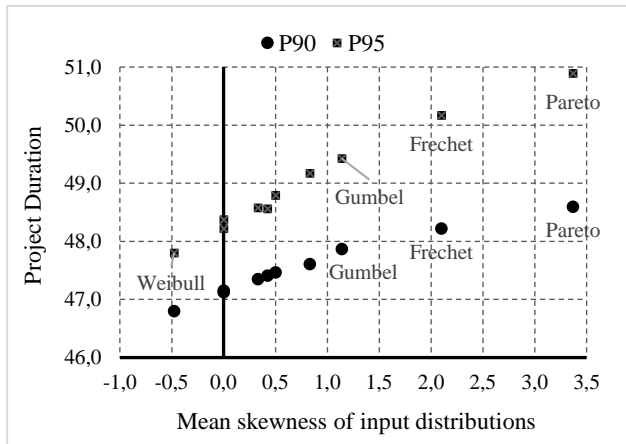
Figure 4. P90 and P95 values of project duration vs. the mean skewness of input distributions

4.5 Effect of Input Distribution Skewness

The relationship between the skewness of the output distribution for the project duration and the mean skewness of the 14 activities was also determined and the result is shown in Figure 5.
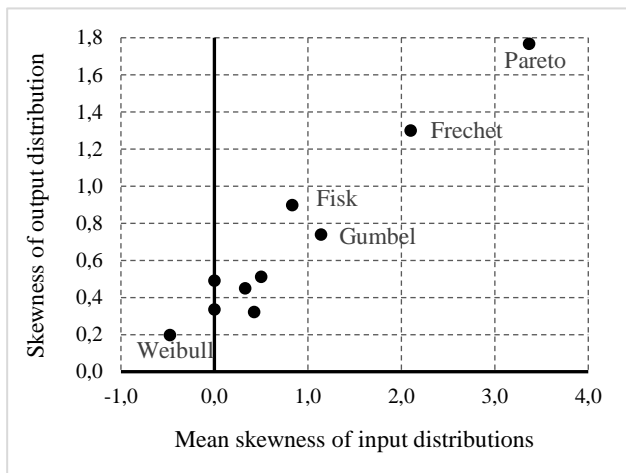
Figure 5. Relationship between mean skewness of input distributions and output distribution skewness

From Figure 5, it is seen that a strong positive correlation exists between the output distribution skewness and the mean skewness of the input distribution for the 14 activities in the network. A correlation value of 0,975 was found for this relationship. Within the range of -0,5 - +3,5 skewness this relationship is close to linear. A straight-line, least squares fit had a regression coefficient of 0,95.

5 Conclusion

The output of a Monte Carlo simulation of a network of activities provides a distribution of the total project duration. The skewness of the output distribution is

positively correlated with the mean skewness of the input distributions. Even if the mean skewness of the input distributions is zero, e.g. for the normal and logistic distribution, the output distribution could be skew. In this study, the skewness of the output distribution is caused by the multiple paths of the project network.

The sum of the mean values of each path is very similar. The data in Table 3 indicates that any of the six paths are possible for one trial of the simulation. This means the critical path varies between the six paths and this causes the skewness of the output distribution, even if the input distributions are all symmetric. It is interesting to note that even for the Weibull distribution, with all activities having a negative skewness, the output distribution has a positive skewness.

The P90 and P95 values of the output distribution are positively correlated with the mean skewness of the input distributions. The correlation values were 0,94 and 0,99 respectively for P90 and P95.

The skewness of the output distribution is strongly correlated with the mean skewness of the activity distributions. The correlation value for this relationship was 0,975 in the range of -0,5 - +3,5. A straight line fit of the data had a regression coefficient of 95%.

The results of this study indicate that the choice of input distribution for cost or schedule simulation should be carefully considered by the risk panel. Total project duration could be underestimated if symmetric or near-symmetric input distributions are selected.

6 Recommendations

In practical applications of Monte Carlo simulation, e.g. in projects, operations or business simulation, a risk panel would decide on which distribution should be used to model the uncertainty in the duration of the activities. Different distributions could be selected for the activities. The next step is to allocate values for the parameters of the distribution that was chosen. The parameters of the triangular or normal distributions are fairly easy to estimate since the scale or location parameters point directly to actual duration. Shankar (2011) commented "the beta distribution can be estimated relatively easily from data on just the optimistic, pessimistic and most likely values". The same applies to the triangular distribution although the mean and variance values differ from the betapert distribution. A stepwise procedure to elicit or estimate values for the triangular distribution is provided by Greenberg (2017).

If actual data for the duration of similar activities in previous projects performed by a company is available, the approach should be to perform a maximum likelihood fit for the data for several distributions and to use the one with the best fit. For novel projects, this data

is seldom available and expert opinion is the only option.

In projects where there is a high degree of certainty in activity duration, e.g. in outage projects for units of a power station, the normal and triangular distributions are recommended for schedule simulations. However, if there is much uncertainty in the duration of some activities in a project, it might be better to use the Fréchet, Fisk (log-logistic) or Gumbel distribution. However, it is not easy to estimate the parameters of these distributions. These distributions have two parameters, typically a shape and scale parameter. The scale parameter relates to the duration of an activity, but the shape parameter is difficult to estimate.

It is therefore recommended that the risk panel start by estimating the parameters of the triangular distribution for all activities, incorporating skewness in the choice of the lower and upper bound values. The mean and standard deviation values of the triangular distributions can then be used to calculate the parameters of the Fréchet, Fisk or Gumbel distribution. These distributions should then be used to run the schedule simulations for a project.

Acknowledgements

References

A. R. Bowden, M. R. Lane, and J. H. Martin. *Triple Bottom Line Risk Management*. John Wiley & Sons Ltd., New York, USA, 2001.

P. J. Christensen and J. Rydberg. overcoming obstacles: strategic risk management in the Øresund bridge project, *PM Network*, 15 (11): 30-36, 2001.

D. Cooper, P. Bosnich, S. Grey, G. Purdy, G. Raymond, P. Walker, and M. Wood. *Project Risk Management Guidelines. Managing Risk with ISO 31000 and IEC 62198*. 2nd Edition, John Wiley & Sons Ltd., UK, 2014.

C. Croarkin and P. Tobias. *e-Handbook of Statistical Methods*. National Institute of Standards and Technology (NIST/SEMATECH), USA, 2012. doi: 10.18434/M32189.

I. Damnjanovic and K. Reinschmidt. *Data Analytics for Engineering and Construction Project Risk Management*. Springer, Switzerland, 2020. doi:10.1007/978-3-030-14251-3.

H. Ehrbar, L. R. Gruber, and A. Sala. *Tunnelling the Gotthard - Gotthard Base Tunnel*. Swiss Tunneling Society, Bauverlag Gütersloh, Switzerland, 2016.

M. Evans, N. Hastings, and B. Peacock. *Statistical Distributions*. 3rd Edition, John Wiley & Sons, New York, 2000.

S. Ferson, L. R. Ginzburg, and H. R. Akçakaya. Whereof one cannot speak: When input distributions are unknown. http://www.ramas.com/whereof.pdf. 1998. (Retrieved 06/07/2020).

M. W. Greenberg. A Step-wise Approach to Elicit Triangular Distribution. *NASA Technical Reports Server*, Bibliogov. 2013.

M. Hajdu and O. Bokor. The effects of different activity distributions on project duration in pert networks, *Procedia - Social and Behavioral Sciences*, 119:766-775, 2014. doi:10.1016/j.sbspro.2014.03.086.

P. E. D. Love, C-P. Sing, X. Wang, Z. Irani, and D. W. Thwala. Overruns in transportation infrastructure projects. *Structure and Infrastructure Engineering*, 10 (2):141-159, 2014. doi:10.1080/15732479.2012.715173.

L. Martens. Schedule Risk Analysis: Case Studies, *Masters Dissertation*. Universiteit Gent, 2017.

N. Munier. *Risk Management for Engineering Projects: Procedures, Methods and Tools*. Springer International Publishing, Switzerland, 2014. doi:10.1007/978-3-319-05251-9.

National Research Council (NRC). *Completing the "Big Dig": Managing the Final Stages of Boston's Central Artery/Tunnel Project*. The National Academies Press. Washington, 2003.

J. M. Nicholas and H. Steyn. *Project Management for Engineering, Business and Technology*. 5th Edition. Routledge, New York, 2017. doi:10.4324/9781315676319.

Project Management Institute (PMI). *A Guide to the Project Management Body of Knowledge (PMBOK Guide)*. 5th Edition, 2013. doi:10.1002/pmj.21345.

Y. Raydugin. *Project Risk Management: Essential Methods for Project Teams and Decision Makers*. John Wiley & Sons Inc., New Jersey, 2013.

W. T. Scherer, T. A. Pomroy, and D. N. Fuller. The triangular density to approximate the normal density. *Reliability Engineering and System Safety*, 82 (3):331–341, 2003. doi:10.1016/j.ress.2003.08.003.

N. R. Shankar, S. S. Babu, Y.I.P. Thorani, and D. Raghuram. Right skewed distribution of activity times in PERT. *International Journal of Engineering Science and Technology*, 3 (4): 2932-2938, 2011.

H. Steyn, M. Carruthers, A. Dekker, Y. Du Plessis, D. Kruger, B. Kuschke, A. Sparrius, S. van Eck, and K. Visser. *Project Management: A Multi-disciplinary Approach*. 4th Edition. FPM Publishing, Pretoria, South Africa, 2016.

Treeplan Software. SimVoi Monte Carlo Simulation Add-in for Excel. https://treeplan.com/simvoi/. (Retrieved 07/06/2020).

M. Vanhoucke. On the use of schedule risk analysis for project management, *The Journal of Modern Project Management*, 2(3):108-117, 2015.

J. K. Visser, Suitability of different probability distributions for performing schedule risk simulations in project management. In Proceedings - *Portland International Conference on Management of Engineering and Technology, PICMET 2016, 4-8 September 2016, Honolulu*, Hawaii, USA, pages 2031-2039, 2016. doi:10.1109/PICMET.2016.7806608.

D. A. Wood, Risk simulation techniques to aid project cost-time planning and management. *Risk Management*, 4(1):41-60, 2002. doi:10.1057/palgrave.rm.8240108.

Development of a model to estimate parameters in a snowpack based on capacitive measurements

Mercedes Noemi Murillo Abril[1]   Beathe Furenes[2]   Nils-Olav Skeie[3]

[1,3] Department of Electrical engineering, IT and Cybernetics, University of South-Eastern Norway, Norway,
mnmurill@gmail.com; Nils-Olav@usn.no
[2] Skagerak Energi, Norway, Beathe.Furenes@skagerakenergi.no

Abstract

In the optimization process of hydropower production, it is relevant to consider some information about the snowpack's parameters. Today, several techniques and devices to measure density, height, and snow water equivalent (SWE) in a snowpack. This paper discusses the development of linear regression models based on voltage measurements collected in a field test of a new measuring device that uses a vertical arrangement of capacitive sensors, to predict density, height, and SWE in a snowpack. The data collected grouped into six data sets and analyzed using the software for multivariable analysis Unscrambler X. From the results, three models were selected, one for each parameter. The models have a good prediction performance within the collection of samples. However, the model data sets used in the process do not have good representativity for other sampling conditions.

Keywords: snow density, capacitive sensor, model development, least square method

1 Introduction

1.1 Background

Hydropower is the main source of energy production in Norway (MPE 2019). Its production can be adjusted through the management of storage water in reservoirs (MPE, 2019). Therefore, having an accurate inflow forecast for the reservoirs optimize the energy´s production. The inflow can vary drastically year to year, but it can be predicted through models that take in count the precipitations of the area. In Norway, approximately 30% of the annual precipitations falls as snow (Saloranta, 2014). The snow precipitations behave as unpredictable buffers of water. Depending on their melting process, they represent less or more water inflow to the reservoirs.

The inflow contribution of the snow buffers can be predicted through models. These models are in constant revision, and one of their sources of improvement is the measurement of snowpack's parameters (Saloranta, 2014). There are several measuring techniques and devices, some of them reviewed in (Kinar and Pomeroy, 2015). The master project reported in (Bjerke et al., 2019) proposes to design a new measuring device based on a vertical arrangement of fixed geometry capacitance plates. This concept, later implemented on a prototype described in a master thesis (Murillo Abril, 2020), required a model to traduce the sensors' measurements into the parameters of interest. These parameters are the snow density by layers, the total height of the snowpack, and its snow water equivalent (SWE).

This project aims to develop and evaluate three different linear regression models, analyze them, and choose the most suitable to calculate each parameter based on the measurements of the prototype.

1.2 Scope and methods

The focus of this paper is the models to calculate the snow density, the height, and the snow water equivalent (SWE) of a snowpack based on the measurements´ values of the prototype, described by Murillo Abril (2020). Due to the warm weather on the zone[1] and mobility restrictions[2], only one data set of samples was collected at the field test. From there, the samples were treated to derivate six data sets for the models´ development.

The multivariate analysis software Unscrambler X 10.3 is used to analyze the data sets. The program has different linear regressions techniques and generates some plots which give a better understanding of the performance of the models. Therefore, the models' selection comes from the analysis of the program's calibration and validation results. The chosen models are simple mathematical relations that can be implemented in software, whether in the prototype´s controller or an external system.

---

[1] There were low precipitations of snow in Porsgrunn area. Its lowest temperatures were around -7ºC (Yr, 2020).

[2] The Norwegian government imposed high fines to people staying outside their municipality (Forbes, 2020), due to the public health emergency of international concern of Covid-19, declared on the 30th of January, 2020 (WHO, 2020).

1.3 Outline of the paper

Section 2 describes the relevant parts of the device used in the data collection process. Section 3 describes the data collection and their posterior treatment. Section 4 shows the data sets' analysis results, calibration, and validation of the models. The results are discussed in section 5, and some conclusions are drawn in Section 6.

2 The device

This section lists the device's features used for data collection during the development of the regression models.

2.1 Concept of measurement

The usage of a vertical arrangement of capacitors as a concept of measurement, discussed in (Bjerke et al., 2019), is based on the relation of the dielectric constant of the snow with its density and the fact that the snow does not accumulate with a uniform thickness (WSL, 2017). Figure 1 shows how the snow has different layers in a snowpack. It is possible to appreciate that the snow looks more compact in the lower layers than at the top. However, there are also intermediate layers more compacted than others on the bottom.
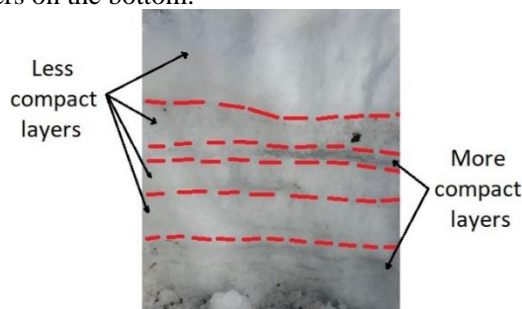
Figure 1: Example of different layers on a snowpack.

Figure 2 shows the sensing system implementing the concept of measuring mentioned earlier. The prototype uses a vertical arrangement of five capacitors installed in fixed distances. The sensor's positions, decided in cooperation with the customer, intent to cover a range of 1.20 meters of snow height with five sensors having a higher accuracy at the lowest layers. The prototype makes it possible to measure the density of up to five layers when the sensors are covered with snow. Because of the snow and air have different dielectric values (Evans, 1965), the prototype's sensors' values in contact with air are different from those covered in the snow. Consequently, the prototype is capable of sensing up to which height the sensors are covered. The SWE is a parameter related to the snow density and height (Brodzik, 2004). Consequently, it is assumed that as the system can recognize the density and height parameters, it can also collect different samples' values in association with the SWE of the snowpack.
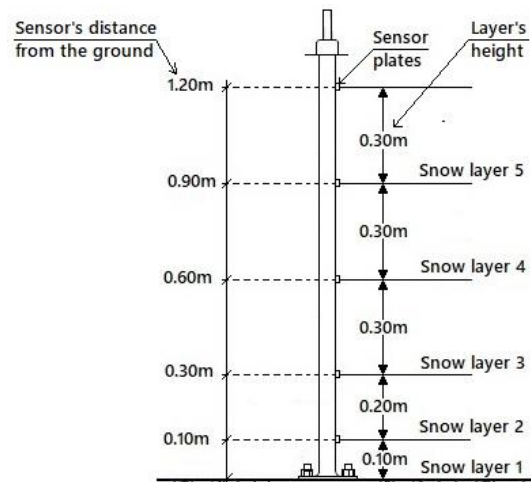
Figure 2: Sensing system of five capacitors.

2.2 The capacitive sensor

The sensor installed in the prototype is the Capacitive Soil Moisture Sensor v1.2. This sensor is typically used in garden applications to sense the humidity in the ground. The sensor has integrated a conducive and ground plates in a conditioning circuit. Figure 3 shows a picture of the installed sensor. The plates are covered by a corrosion-resistant material (How to Electronics, 2019), and act as capacitors of fixed geometry, varying their capacitance when surrounded by different mediums. The conditioning circuit uses a fixed frequency oscillator that feeds the plates with a square signal, and depending on its capacitance, the output signal's voltage varies (How to Electronics, 2019). For this prototype, the sensors are powered with 5 Volts DC, and their outputs are connected to the analog inputs of the Arduino Nano V3.0.

Figure 3: Capacitive sensor installed in the prototype (SwitchDoc, 2020).

The Arduino Nano is a small and complete board based on a microcontroller (Arduino, 2008). The microcontroller has eight analog inputs with an analog-digital converter (ADC) of 10 bits resolution (Circuits Today, 2020). The prototype uses the first five analog inputs to read the sensor's signals from 0 to 5 volts with a change of 4.88mV.

2.3 Construction and behavior

The prototype, shown in Figure 4, uses drainpipes tubes and couplings to have a modular hollow structure. This structure allowed to place the sensors distanced, as shown in Figure 2 Inside the structure are the electronic circuits of the system, and the used material prevents the

leaks of water when the device is covered by snow. On top of the sensors' column is the Arduino board placed.
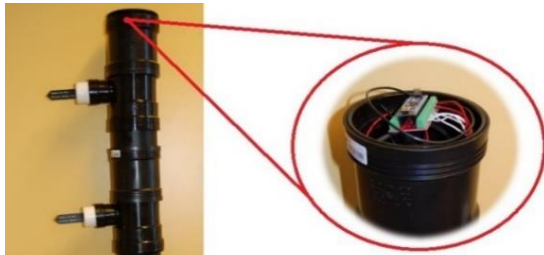
Figure 4: Prototype with five sensors at different heights and controller board on the top (Murillo Abril, 2020).

Figure 5 shows the flow of information on the system. The sensor senses the environment and conditions the output analog signals, as explained in the previous section. The microcontroller receives the signals through the ADC, processes and sends them through the serial port, incorporated on the board. To control the information flow in the prototype, the microcontroller's software has implemented an algorithm shown on the state diagram in Figure 6**.** The algorithm is designed to continuously read the system's sensors' values from the bottom to the top.
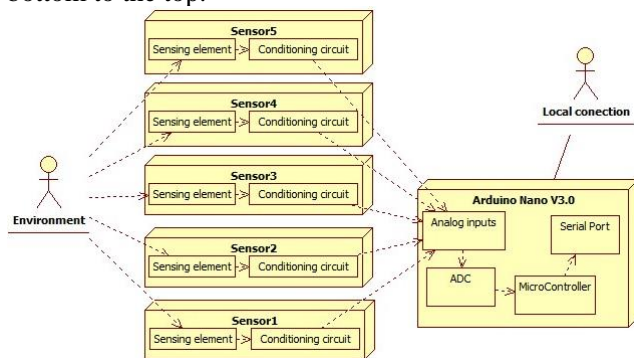
Figure 5: Diagram of the information flow through the prototype.

When the data collection process starts, the first state of the algorithm initializes a counter that identifies the sensor. It always begins from the lowest sensor first. In the next state, it reads the value of the sensor. Here a string variable is updated with the timestamp of the capture, the sensorID (sen+sensor number), and the captured value. The next state prints this variable in the serial port to be read in a serial monitor through a local connection.
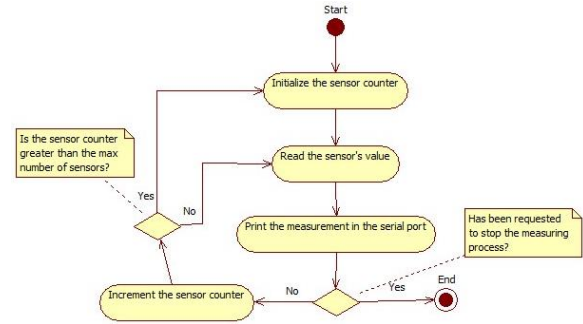
Figure 6: State diagram for the measuring process.

3 Modelling data sets

This section explains the collection and posterior treatment of the modeling data.

3.1 Data collection

The data collection took place in Viddaseter, one of the ski areas in Greenland (Visit Telemark, 2019). The location is about 500 hundred meters over the sea level. The field trip to collect the data was on the 26 of March 2020. The temperature of the day was 5.5ºC. The first step was to obtain a manual sample of snow density and height. The sample was taken using a cylindrical gauge and a marked metal stick, respectively. Around the manual sampling place, the terrain was prepared to dig a rectangle hole to have enough space for the prototype's installation, see Figure 7. The values of the density and height manually measured were 389.864 Kg/m^3 and 0.60 meters, respectively.

Figure 7: Preparation of the area to take the manual sample and later installation of the prototype.

In order to have more samples to train the model, the prototype was placed in two different places of the same pit. Figure 8 (left) shows the device placed on the first measuring place. Here the device was introduced on the snow. The first two sensors were fully covered on the snow, the third sensor was close to the top of the snow and the two remaining sensors were just surrounded by air. The second measuring place was in the opposite part of the pit. Figure 8(right) shows the device placed on the second point. Here the device was covered fully with snow up to the third sensor. The two remaining sensors on the top were surrounded by air.

Figure 8: First measuring place (left), second measuring place (right).

To record the sensor's measurements' values, the system's controller had a local serial connection with an external computer, see Figure 9. The laptop had running the open software Tera Term, which is a serial terminal emulator. The emulator records the readings of the computer's serial port in a buffer. The measuring process started once the prototype was placed and stable in the measuring position. The process was run seven times. The first three tests were in the first measuring place. The tests 4, 5 and 7 in the second measuring place and test 6 all the sensors surrounded only by air. The mediums around the sensors during the tests are listed on Table 1. After each run the information on the buffer of Tera Term was saved into to a comma-separated values file.

Figure 9: Data collection using an external computer in serial communication with the prototype.

Table 1: Conditions of the tests.

Test \ Sensor	Sen1	Sen2	Sen3	Sen4	Sen5
1	snow	snow	air	air	air
2	snow	snow	air	air	air
3	snow	snow	air	air	air
4	snow	snow	snow	air	air
5	snow	snow	snow	air	air
6	air	air	air	air	air
7	snow	air	air	air	air

3.2 Data treatment

In a linear regression model, it is necessary to have a set of matched samples for the independent and dependent variables of the problem (Esbensen, 2010). The samples are used to calculate regression coefficients, which complete the general equation regression (Esbensen, 2010), presented in equation (1). In univariate regressions, the samples are divided into two matrices of one column, one containing all the values of the independent variable (x) and the second one with all the dependent variable's values (y).

$$y = b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n \qquad (1)$$

In one case of the multivariate regressions, there is only one dependent variable, but there is more than one independent variable. In this case, the values of the independent variable are arranged on a matrix (X) where each column corresponds to the values of each variable while the matrix of the dependent variable still is of one column. The values recorded during the field test of the prototype constitute raw measurements of the sensor's voltage. These are the independent variables of the regression, which has been arranged differently. The next three parts detail the arranging process.

3.2.1 Pretreatment

The samples from tests 1, 2, and 4 were not used for further analysis. They were measured using an old version of the controller's program, that did not work correctly. Before the regression matrices were constructed, the averages of the samples for each test were plotted, see Figure 10. The standard error (Francis, 2018) of the samples used to construct Figure 10 is 0.001 volts. It is possible to identify that the Test 3 values did not match the conditions of measurement stated in the third row of Table 1. In this test, the average for the sensors Sen1 and Sen2 were expected to be close to the ones in Test 5, where these sensors are also covered with snow. However, their values are closer to the averages in Test 6, where these sensors are exposed to air. Consequently, the only values used to develop the models were the ones from test 5 to 7.
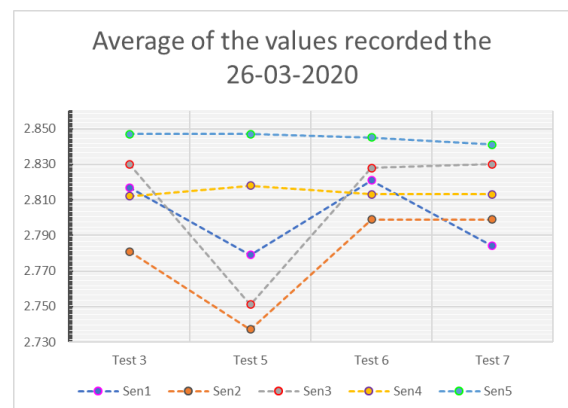
Figure 10: Average of the values by test run 3, 5, 6 and 7.

3.2.2 Snow density

According to the measuring concept, each sensor measures the density related to the material, which is in their surroundings. Therefore, the model to predict the snow density (dependent variable) requires the voltage value (independent variable) of the sensor in contact with the snow, as shown in equation (2). The X matrix consists of the record voltages of the sensors, from the tests mentioned before, stack on a column. There was only one manual sample for the density of the snow. Consequently, the Y matrix was completed with two o different values. When the sensor was covered by snow, the inserted value was the manual sample (389.864 Kg/m3). When the sensor was only surrounded by air, the inserted value was 1.191 Kg/m3, which corresponds to the air's density at a pressure of 95276.7 Pa and a temperature of 5.5 °C (Czenia and Haponiuk, 2020).

$$\rho_{snow} = b_0 - b_1 V_{sen} \qquad (2)$$

These two columns of values form the Raw Data Set (SD). Applying a moving average filter to the measurements of the Raw Data Set (SD), two additional data set were created. The construction of the FilterX4 Data Set (SD) is shown in Figure 11. There the filter uses four values from the raw measurements. The same notion was used to create the FilterX8 Data Set (SD), but the filter used eight raw measurements. To the validation of the models, the technique used was the cross-validation. The constructed datasets are then divided into two parts. The first part used to calibrate the models and the second to validate them.

Figure 11: Construction of the FilterX4 Data Set (SD) from the Raw Data Set (SD) for calibration and validation of the Snow density model.

3.2.3 Total height and SWE

For the height and SWE parameters, the measurement concept is related to the number of sensors covered with snow during the sampling. Therefore, the construction of the matrix X for both cases was the same. This time, the sensors' values were independent variables, which their combination of values varies depending on the height and SWE of the surrounding. For the Y matrix, the inserted values vary depending on the parameter of interest. For the total height, the inserted values were corresponding to the measured height in the field. For the SWE, the inserted value depended on the number of

layers' height cover by snow. The inserted value was calculated using the equation (3) (Murillo Abril, 2020). There, the terms of 'z' represent the height of the covered layers in meters. Consequently, the SWE values for tests 5, 6, and 7 were completed with 233.92; 0.00; and 38.99 meters.

$$SWE = (z_1 + z_2 + \cdots + z_n)\frac{389.864}{100} \qquad (3)$$

Figure 12 partially shows the X and Y matrices used to form the height and SWE data sets. As in the previous part, the tests' measurements were stack to build a Raw Data Set (H) and (SWE) for the height and SWE, respectively. Using the same filter technique, as for the density, over the values of each Raw Data Set created additional data set to analyze. For the height, the data sets used were Raw Data Set (H), FilterX4 Data Set (H), and FilterX8 Data Set(H). While for the SWE, the data sets analyzed were Raw Data Set (SWE), FilterX4 Data Set (SWE) and FilterX8 Data Set (SWE). As in the previous section, the data sets are divided into two parts to the calibration and validation of the models.

Figure 12: Partial X and Y matrices of raw data for height and SWE data sets.

4 Results

This section reports the results obtained from the calibration and validation processes of the models generated with the Unscrambler X 10.3.

4.1 Snow density

The snow density regression model corresponds to the analysis of univariate samples; therefore, it is only valid to calculate the snow's density by layers. In Unscrambler X, the multivariate linear regression (MLR) was performed over the data sets but keeping the X matrix with only one column. This software then used the Least Square Regression (LS-R) method to minimize the total sum square of errors (SST) related to the data set. The SST value is calculated adding the square errors between the samples and a regression model as the average of the samples (Flotz, 2013). After finding the parameters of the calibrated models, the square sum of errors of the samples against the new regression models are calculated (SSE). The first two rows of Table 2 have the SST and SSE from the Unscrambler X analysis. Table 2 last row is the error reduction in percentage, and

it was calculated from the ratio of the SSE and SST for each data set.

Table 2: SST, SSE and Error reduction (%) for the Raw Data Set (SD), FilterX4 Data Set (SD) and FilterX8 Data Set (SD).

Data Set Error	Raw (SD)	FilterX4 (SD)	FilterX8 (SD)
SST	16783250	14037500	10353740
SSE	5083899	3462451	2489117
Error reduction	30%	25%	24%

In all the cases, the SSE is reduced, so to the select one over the other two models, it was calculated the ration of reduction. In the same order as in Table 2, the errors were reduced to 30%, 25%, and 24% of the original values. The model chosen was from the FilterX8 data set for reducing the most from the original error. The calibrated model is presented on the equation (4). There the ρ_{snow} is the density in Kg/m3 of the material in the surrounding of the sensor and V_{sen} corresponds to the voltage measure by the sensor in Volts.

$$\rho_{snow} = 12565.58 - 4437.733 V_{sen} \qquad (4)$$

The model's performance to fairly predicts the density with new measurements was evaluated using the Predicted vs. Reference plot. The plot corresponding to the selected model is in Figure 13. The blue data corresponds to the evaluation using the calibration data set while the red corresponds to the evaluation using the reserved data set for the validation. The statistics, shown in the upper left corner of the plot, indicates that using the data that does not intervene in the model's calibration has the closest slope to the perfect prediction (1). The RMSE is high for both cases, but with the validation set is around 6 points lower.
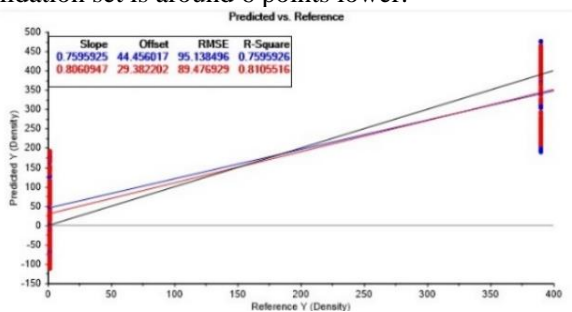
Figure 13: Predicted vs. Reference plot for the calibration and validation of the model from the FilterX8 Data Set (SD).

4.2 Total height

The samples for the height model are multivariate data sets. The Partial Least Square Regression (PLS-R) was run the Unscrambler X 10.3. The software decomposes the X matrix generating several models using the decomposed factors for each data set and automatically choose the one that has explained the variance of the

data set with the lowest numbers of factors. In this case, in the analysis of each data set, the software determined two-factor models. Manually the plots of Root Mean Square Error (RMSE) explained against the factors used in the results were examined to corroborate the software determination. To select one of the models, it was analyzed the Predicted vs. Reference plots for each data set. These plots are in Figure 14, Figure 15, and Figure 16, respectively. From those, the model from the FilterX8 data set has the lowest RMSE. Consequently, it is the chosen model, and its equation is (5). From this expression the h represents the height of the snowpack in meters and the V terms are the voltage values of the sensors in Volts.

$$h = 2295.096 - 621.189 V_{sen1} - 326.097 V_{sen2} \\ - 244.805 V_{sen3} + 311.173 V_{sen4} \qquad (5) \\ + 62.12442 V_{sen5}$$

As in the previous part, the Predicted vs. Reference plot also evaluates the models' performance to predict the height from new voltages' values. The chosen model has a slope of 0.97, only three hundredths apart from the unit.
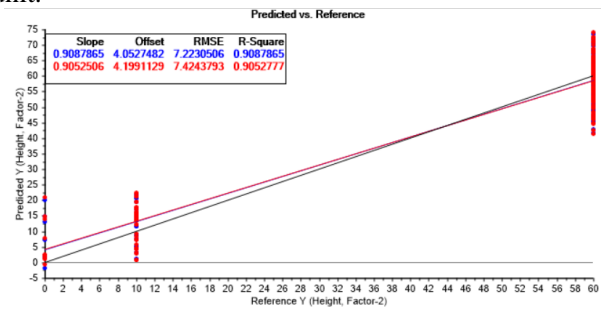
Figure 14: Predicted vs. Reference plot for the calibration and validation of the model from the Raw Data Set (H).
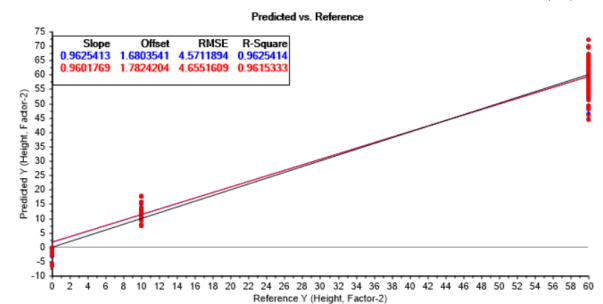
Figure 15: Predicted vs. Reference plot for the calibration and validation of the model from the FilterX4 Data Set (H).
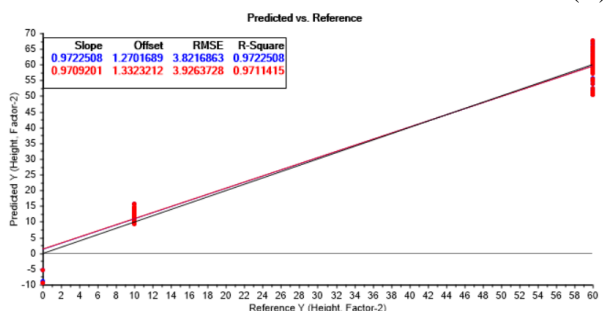
Figure 16: Predicted vs. Reference plot for the calibration and validation of the model from the FilterX8 Data Set (H).

4.3 Snow water equivalent

For the SWE model, the procedure used for the implementation in the height model was executed. The Predicted vs. Reference plots of the data sets are in Figure 17, Figure 18, and Figure 19, respectively. From the plots, the lowest RMSE corresponds to the FilterX8 data set. This is the model chosen and its equation is written in (6). The SWE is the parameter of interest given in millimeters and the V terms are the voltage values of the sensors in Volts.

$$\begin{aligned} SWE = 8947.771 &- 2421.820 V_{sen1} \\ &- 1271.320 V_{sen2} - 954.394 V_{sen3} \\ &+ 1213.319 V_{sen4} \\ &+ 242.2006 V_{sen5} \end{aligned} \quad (6)$$

The model's performance to predict the parameter from new measurement values, given by the slope in Figure 19, is close to the unit.
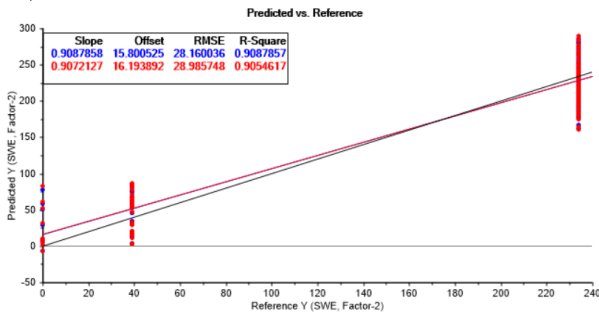
Figure 17: Predicted vs. Reference plot for the calibration and validation of the model from the Raw Data Set (SWE).
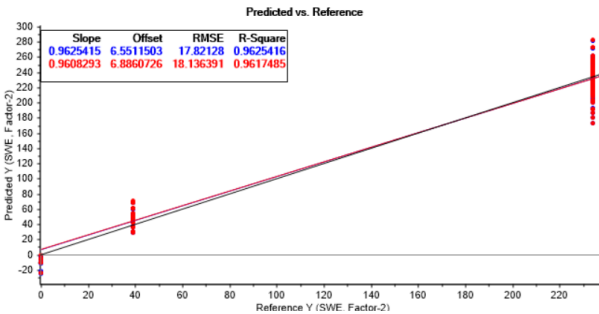
Figure 18: Predicted vs. Reference plot for the calibration and validation of the model from the FilterX4 Data Set (SWE).
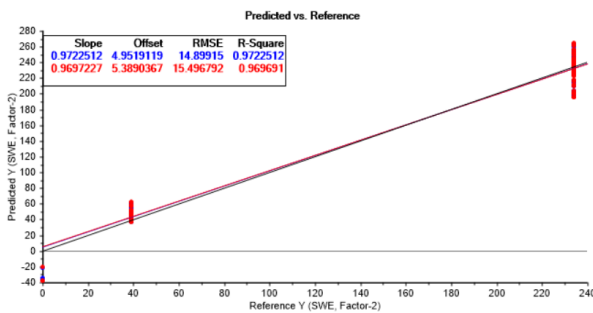
Figure 19: Predicted vs. Reference plot for the calibration and validation of the model from the FilterX8 Data Set (SWE).

A relevant result to include is the plot of the X-Loading Weights, see Figure 20. In the analysis of the height and SWE models, the X-Loading Weights plot gave the same results. It shows a higher negative contribution from the values of the sensors 1 to 3. The two remaining sensors have a lower positive contribution. This might be a consequence of the variation of snow coverage during the sampling process only on the three first sensors.
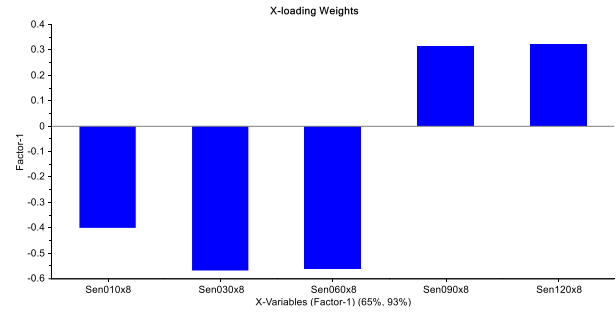
Figure 20: X-loading Weights plot for the analysis of the height and SWE models.

5 Discussion

In the development of regression models is important to have a rich data set of samples. Those must include as much information as possible in different measuring scenarios varying places, weather conditions, etc. However, due to the weather and mobility restrictions, the samples were collected in a single place without other variable conditions. Additionally, four of the seven test measurements were discarded due to errors during the measuring process. These two conditions make that the resulting models do not have a good prediction performance over other samplings scenarios.

In the models, the ones resulted from the FilterX8 Data Sets show better performance against the others. This can be the result that when the samples are filtered, noise errors are reduced in the data set. It is important to remark that the difference in performance between the models of the FilterX4 and FilterX8 data sets is not large. It is possible that increasing the values to use in the filter will increase the models' performance, but not a significant amount. Further works can test a model using 12 values in the filter to corroborate or discard the hypothesis. Also, it should analyze if the resources used to take more measurements and processes them contributes significantly to support a decision.

The slopes close to one in the Predicted vs. Reference plots indicate that the three chosen models' performances in the prediction of the parameters are close to the perfect performance. However, the confidence in these results cannot be high. The method used for the validation of the model was the cross-validation. Even the validation of the models was done with a separate data set that did not intervene in the calibration process. It comes from a single sampling data set. It means that the models might be representing

the errors associated with the sampling process more than the actual variance in the parameters of interest. Further work can include a new sampling data set to validate the models and revalidate its prediction performance.

The X-Loading Weights of the height and SWE models indicate that the principle of measurement can capture representative data to calculate the parameters of interest. A better structure of the sampling process can corroborate this theory and improve the models. A suggestion for the sampling process can include a step increase of the height of snow covering the sensors adding one at the time. For further works, the deployment of more than one device in different locations can increase the number of samples, and it can also improve the representativity of the data sets for calibrating and validating the models.

6 Conclusion

The Predicted vs. Reference plots in Figure 13, Figure 16, and Figure 19 show that the models generated have a good performance to predict the parameters of interest in the snowpack. However, this is only within the samples collected. The variance of the sensors' snow coverage during the sampling process, has a high impact in the loadings of the sensors' values for the height and SWE models. A manual measurement for the density that matches the positions of the sensors might introduce more important information to the sampling data sets.

References

Arduino. *ARDUINO NANO: Overview*. 2008. [Online]. Available: https://store.arduino.cc/arduino-nano. [Accessed 23 June 2020].

J. M. Bjerke, M. N. Murillo Abril, A. Jaganjac, and N. Pouladi. *Measurement of snow density*. Master's Project. University of South-Eastern Norway, Porsgrunn, Norway, 2019.

M. J. Brodzik. *The relationship of snow water equivalent to snow depth and density*. 2004. [Online]. Available: http://webarchiv.ethz.ch/arolla/Arolla_Data/SnowConditions/depth_to_swe.pdf. [Accessed 23 June 2020].

Circuits Today. *Arduino Nano Tutorial – Pinout & Schematics*. 2020. [Online]. Available: https://www.circuitstoday.com/arduino-nano-tutorial-pinout-schematics. [Accessed 23 June 2020].

D. Czernia and B. Haponiuk. *Air Density Calculator*. 2020. [Online]. Available: https://www.omnicalculator.com/physics/air-density. [Accessed 20 May 2020].

K. H. Esbensen. Multivariate Calibration (PCR/PLS). in Dominique Guyot, Frank Westad, and Lars P. Houmøller, editors, *Multivariate data analysis in practice*, pages 115-154. CAMO Software AS, 2010. ISBN 82-993330-3-2.

S. Evans. Dielectric Properties of Ice and Snow–a Review. *Journal of Glaciology*, 5(42): 773-792, 1965. doi: 10.3189/S0022143000018840.

B. Flotz. *Statistics 101: Linear Regression, the Very Basics*. 2013. [Online]. Available:

https://www.youtube.com/watch?v=ZkjP5RJLQF4. [Accessed 21 May 2020].

Forbes. *Norway Hands Out $2,000 Fines Or Jail For Ignoring Coronavirus Quarantine*. 2020. [Online]. Available: https://www.forbes.com/sites/davidnikel/2020/03/17/norway-hands-out-2000-fines-or-jail-for-ignoring-coronavirus-quarantine/#1bedeaf44f42. [Accessed 2 July 2020].

P. Francis. *Standard Uncertainty, Standard error and Standard deviation*. 2018. [Online]. Available: https://www.youtube.com/watch?v=NzrOfAsgxe8. [Accessed 04 September 2020].

How to Electronics. *Interface Capacitive Soil Moisture Sensor v1.2 with Arduino*. 2019. [Online]. Available: https://how2electronics.com/interface-capacitive-soil-moisture-sensor-arduino/#Capacitive_Soil_Moisture_Sensor_v12. [Accessed 19 May 2020].

N. J. Kinar and J. Pomeroy. Measurement of the physical properties of the snowpack. *Reviews of Geophysics*, 53(2): 481–544, 2015. doi: 10.1002/2015RG000481.

Ministry of Petroleum and Energy. *Electricity Production*. 2019. [Online]. Available: https://energifaktanorge.no/en/norsk-energiforsyning/kraftproduksjon/. [Accessed 20 February 2020].

M. N. Murillo Abril. *Development of a remote measurement*. Master's thesis. University of South-Eastern Norway, Porsgrunn, Norway, 2020.

T. Saloranta. *Report no. 6 – 2014, New version (v.1.1.1) of the seNorge snow model and snow maps for Norway*. Norwegian Water Resources and Energy Directorate, Oslo, Norway 2014.

Swiss Federal Institute for Forest, Snow and Landscape Research WSL. *Snowpack*. 2017. [Online]. Available: https://www.slf.ch/en. [Accessed 21 February 2020].

SwitchDoc. *Tutorial: Waterproofing your Capacitive Moisture Sensors*. 2020. [Online]. Available: https://www.switchdoc.com/2020/07/tutorial-waterproofing-capacitive-moisture-sensors/. [Accessed 4 September 2020].

Visit Telemark. *Viddaseter*. 2019. [Online]. Available: https://www.visittelemark.no/ting-a-gjore/viddaseter-p502523. [Accessed 23 June 2020].

WHO. *Rolling updates on coronavirus disease (COVID-19)*. 2020. [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen. [Accessed 10 May 2020].

Yr. *Statistics – Porsgrunn*. 2020. [Online]. Available: https://www.yr.no/en/statistics/graph/1-31680/Norway/Vestfold%20og%20Telemark/Porsgrunn/Porsgrunn?q=2020. [Accessed 2 July 2020].

Real-time monitoring of wood cladding spray painting properties and nozzle condition using acoustic chemometrics

Maths Halstensen[1]   Onyinye Victoria Agu[1]   Ulrich Hundhausen[2]   Kim H. Esbensen[3]

[1] Department of Electrical Engineering, IT and Cybernetics, University of South-Eastern Norway, Norway, maths.halstensen@usn.no
[2] Norwegian Institute of Wood Technology (NTI), Oslo, Norway, uhun@treteknisk.no
[3] KHE Consulting, Copenhagen, Denmark, khe.consult@gmail.com

Abstract

An experimental setup simulating real-time wood cladding coating monitoring of nozzle conditions and spray paint properties has been investigated. This approach is based on affixed nozzle accelerometer sensors with appropriate signal conditioning and chemometric data analysis (PCA). The aim was to develop effective visualization of different process states using PCA score plots. The present feasibility study shows that this approach can be used as a basis for further development towards a Process Analytical Technology (PAT) spray monitoring system able to work in the harsh environment of an industrial wood cladding paint box. However, there is still a significant amount of on-site industrial calibration and R&D necessary before a final method validation can be executed. The present results rely on permanently affixed PAT sensors. Further studies will a.o. focus on the degree to which replacement of acoustic accelerometer sensors necessitates recalibration of the multivariate data models employed, which is a critical success factor in industrial implementations.

Keywords: process analytical technology, PAT, multivariate data analysis, accelerometer, principal component analysis, process monitoring, nozzle spray condition, statistical quality control

1 Introduction

Wood has been the common façade material in Scandinavia for centuries and industrial application of exterior coatings has been markedly increasing. While quality control in the cladding industry is still mainly carried out manually, automated in-line measurements are on the rise to secure greater consistency and improve quality and productivity (Hundhausen et al., 2016).

Development of an on- and in-line quality control system at one of Europe's largest coating plants for cladding is addressed in an ongoing R&D project termed "KonTre" (Kontre). By analysis of data from many sensors located at critical stages in the full production line, necessary input/output data can be obtained for multivariate calibration and model predictions, which are essential for continuous process and product optimization. The control of the coating application is of overall importance as the film thickness is positively correlated to the product quality in terms of the coating's service life and, thus, the maintenance interval of a building façade (Grüll et al., 2014).

Regarding process yield, an instantaneous detection of spraying defects is crucial particularly in view of high feed speeds of up to 200 m/min in industrial coating of cladding (Hundhausen et al., 2018). While abrupt failures like full clogging are easy to detect, gradual changes like nozzle wear, i.e. erosion of nozzle orifices, are not.

Acoustic chemometrics (Halstensen et al., 1998; 2000; 2001; 2006; 2010; Esbensen et al., 1998; 1999; Bakeev et al., 2010; Ihunegbo et al., 2012; Arvoh et al., 2012; Wagner et al., 2013) has proven to be a powerful tool for monitoring nozzles in various applications, such as characterization of liquid flow through an orifice plate nozzle. Also monitoring viscosity of anti-icing fluid used on aircrafts in below zero conditions has shown promising results (Halstensen et al., 2019). Its usability in wood coating processes has however not been investigated yet.

The present study aimed to assess the feasibility of using an acoustic chemometrics approach to monitor the spray nozzle performance in industrial coating of wood cladding. This was done by addressing following research questions:

1) Will the acoustic chemometrics approach, which includes Principal Component Analysis (PCA) (Esbensen and Swarbrick, 2018) and eventually Partial Least Squares regression (PLS-R), be able to extract useful information about nozzle orifice wear and clogging?
2) Will the signal from the sensor still be representative for the spraying nozzle performance or need re-calibration after replacing a worn-out nozzle after typically ∼1000 l of paint spraying?
3) Will potential build-up of paint on the spraying nozzle and sensor affect the measurement adversely?
4) Will the method be sensitive to the paint type or can a global multivariate model be used to monitor the process independently of the paint type?
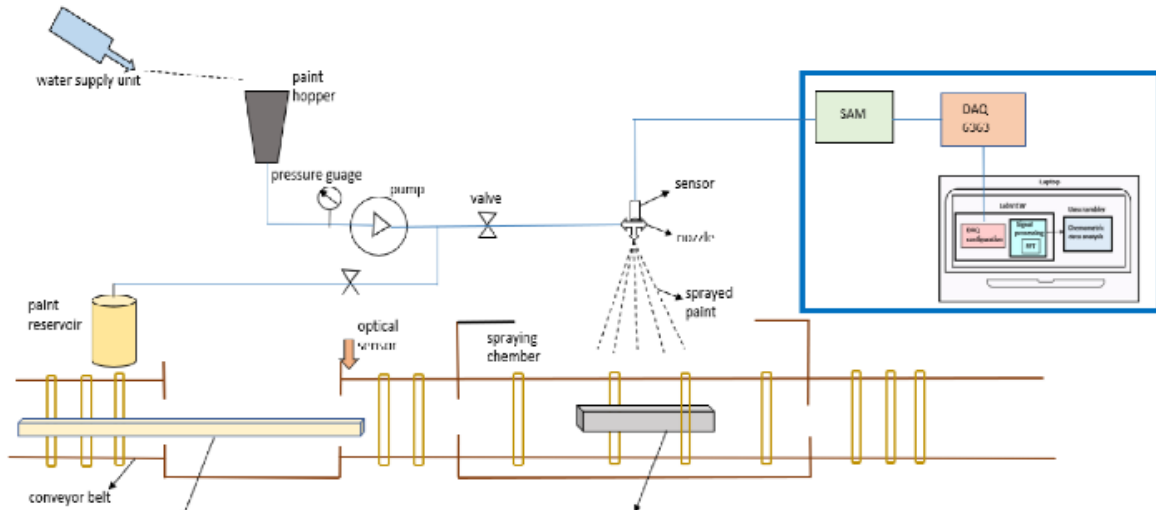
Figure 1. Simplified illustration of the experimental spraying test rig equipped with an acoustic PAT-sensor.

2 Materials & methods

2.1 Equipment

Industrial conditions were simulated using an experimental coating line in the laboratory of the coating manufacturer Jotun AS in Sandefjord, Norway (Figure 1). The test rig consisted of a conveyer belt for the wood, an optical sensor that activates spraying, a spray chamber with an airless spraying nozzle, a water supply unit, paint hopper, pressure gauge, pump, paint mixer, paint cooling unit, hose pipe, and a handheld thermometer. The installed PAT hardware was an acoustic sensor (accelerometer) Bruel and Kjær model 4519, Signal Amplifier Module (SAM) developed at the University of South-Eastern Norway, a data acquisition device USB-6363 from National Instruments and a standard laptop computer. The software used was LabVIEW (version 2019, NI, Austin, Texas, USA), Unscrambler X (version 10.3, Camo Analytics AS, Oslo, Norway).

2.2 Experimental design

A full factorial design (Esbensen and Swarbrick, 2018) was employed comprising 64 experiments:

$$h^n \cdot a = n \tag{1}$$

where h (levels) = 2, n (factors) = 4, a (paints) = 4 and n = the number of experiments.

The four different paints from Jotun AS (Sandefjord, Norway) were water emulsions based on two different binders, Jotun Industri Opaque Primer and Jotun Industri Grunning Visir, and the topcoat Jotun Industri

Optimal in two different bases (
Table 1). The paints of Base A are bright due to a high content of titanium dioxide (TiO_2), those of Base C are dark.

Table 1. The four different paints used in the experimental design.

Paint Name	Type	Binder	Base	Titanium content [%]	Solid content [vol%]	Specific gravity
Opaque(1)	Primer	alkyd	A	16	46±2	1.45
Visir(2)	Primer	alkyd	C	0	36±2	1.04
Optimal(3)	Topcoat	acrylic	A	16	40±2	1.25
Optimal(4)	Topcoat	acrylic	C	0	40±2	1.25

The four factors were 1) paint temperature, 2) pump pressure (paint flowrate), 3) nozzle size, and 4) nozzle condition in terms of partly clogged or free flow.

The influence of the paint temperature on the acoustic signal was tested on two levels (Table 2). The temperature levels were measured with a handheld infrared thermometer.

Table 2. Levels for the factor temperature of the four different paints in

Table **1**.

Paint type	High	Low
Paint 1	21 ºC	8 ºC
Paint 2	20.8 ºC	9.6 ºC
Paint 3	20 ºC	7.1 ºC
Paint 4	20 ºC	7.5 ºC

The specific pump pressures (factor 2) applied for each paint were 0.4 and 0.3 bar. The orifice diameter of the nozzles (factor 3) were 0.025 inch (0.635 mm) and 0.019 inch (0.483 mm), respectively. The two nozzle sizes (factor 3) were tested under clogged and open conditions while other variables remained constant (high and low temperature, high and low pressure). Both nozzles had a spraying angle of 60°.

The influence of the nozzle condition (factor 4) on the acoustic signal was studied on nozzles with a free flow and nozzles that were partly clogged. The latter was induced by forcefully inserting a wedge-shaped wood splinter in the nozzle opening (Figure 2). As this was done manually, the clogging conditions were not identical in all experiments. In the eleventh measurement, the clogging material fell out towards the last part of the experiment, conveniently simulating a transient clogging situation.
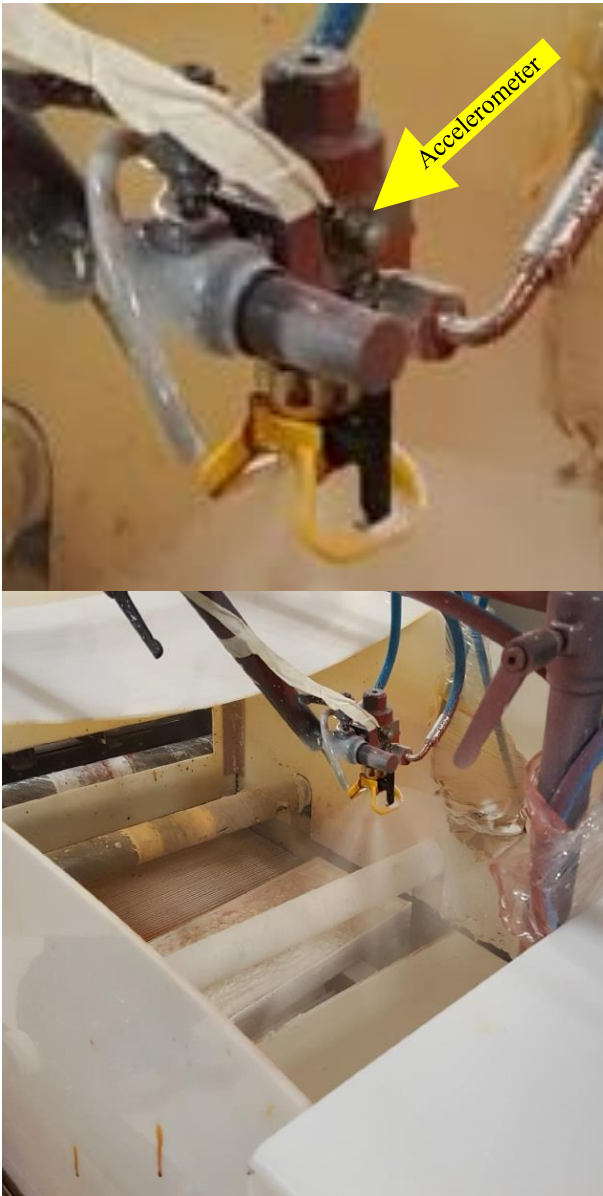
Figure 2. Upper: Spraying nozzle assembly with accelerometer. Lower: Disturbed spray pattern from a partly clogged nozzle.

Each paint type was mixed thoroughly in the paint mixer to achieve homogeneity. The paint was transferred to the hopper, ready to be pumped through the rig (Figure 1). Before applying a new paint (Table 1), hosepipes a.o. parts were rinsed with high-pressure water.

The temperature of the paint was measured manually with a handheld thermometer before it was pumped into the hopper in the test rig.

Once the optical sensor mounted on the conveyor belt detected a wood board, the nozzle started to spray paint, and the acoustic sensor started to capture the measurement signal in each test, 100 acoustic spectra were captured and then converted to one spectrum to get a precise measurement. This was repeated 5 times to give 5 replicates, which was saved on the computer. At the end of each experimental run, the used paint was transferred into the waste container.

2.3 Acoustic chemometrics

Figure 3 shows the acoustic chemometric signal path involved in acquiring vibrational signals from the active nozzle and their subsequent signal processing.
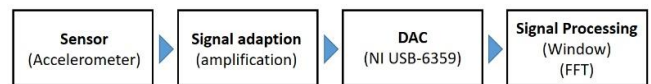
Figure 3. Block diagram of processes involved in acquiring vibrational signal from active spraying nozzles.

The acoustic sensor used is a piezoelectric shear accelerometer type 4518. In general, this type of PAT sensor is used for vibration monitoring, modal and structural analyses. Four different piezoelectric accelerometers were used for different experimental tests, which are discussed in the later sections. The piezoelectric sensor generates an electric charge signal proportional to the vibration acceleration due to the flow of paint through the nozzle bore and its sudden depressurized spraying action upon leaving the bore orifice.

The electric signal is produced when the inbuilt crystal material is subjected to a shear force. When the mass attached on the crystal material experiences a vibration, it causes the piezoelectric crystal to stretch and rebound, thereby generating opposite charges that are transmitted as an electric signal. The piezoelectric sensor has an inbuilt amplifier that converts the output charges into a low-impedance voltage signal between 1 and 100 mV. The sensor has a high resistance to base bending and a high response frequency range > 20 Hz, and can also withstand temperature changes between -51 and 121°C. Its low impedance output characteristic leads to low operating voltage, making it more economical and safer for use. It has low resistivity, i.e. any small signal can be detected and transmitted. Its output impedance < 100 Ω. The other end of the sensor cable was connected to the input channel port on the Signal Amplifier Module (SAM).

The SAM amplifies the signal from the acoustic sensor (1 – 100 mV) in order to use the entire input range which is ±10 V. Having a large signal range will guarantee maximum digital resolution in the following

analog-to-digital converter. When the signal is amplified, the signal amplitude obtained is about ±5 V, which will then flow to the DAQ unit.

The USB 6363 data acquisition (DAQ) data acquisition hardware is used to acquire a signal (like analog, digital, etc.) from a sensor into a laptop. It is capable of handling analog to digital signal conversion, A/D. Since the signal from SAM is accessed in the computer, the SAM signals must be converted from analog to digital format.

The DAQ 6363 has a 16 bits resolution (i.e. the smallest change the unit can measure). Its maximum sampling frequency (how fast sampling is done) rate is 2 MHz. The vibration frequency is between $0 - 100$ kHz. Thus, the sampling rate used in this work is 200 kHz to satisfy Shannon's sampling theorem. The configuration of this unit was carried out in LabVIEW.

The most important signal processing steps involved is shown in Figure 4, where a time series of 4096 samples corresponding to 0.02 sec is recorded from the accelerometer. The time series is then multiplied with a Blackman Harris window (Ifeachor and Jervis 1993) canceling out the signal towards the ends of the series to prevent spectral leakage. The final step is the Discrete Fourier Transform (DFT) that is used to transform the time domain signal into the frequency domain. The DFT can be expressed as

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi kn/N} \quad k = 0, \dots, N-1 \qquad (2)$$

A more efficient implementation of the DFT is the Fast Fourier Transform (FFT) (Ifeachor and Jervis, 1993) which in this work has been implemented in LabVIEW 2017 for fast real time calculation of the Fourier spectrum. Figure 4 shows a part of the LabVIEW GUI where the time series (4096 samples) is shown on top, and the FFT spectrum below. This spectrum represents the vibration signal in the pertinent time interval; it can be viewed as a "spectral fingerprint" of the state of the Normal Operating Conditions (NOC) of the nozzle while at work.
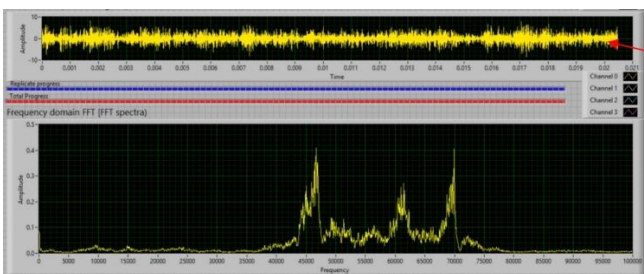
Figure 4. Top: time domain signal indicated with a red arrow, Bottom: derived frequency domain (FFT) spectrum. This spectral fingerprint is used directly in the multivariate calibration of a state prediction model (see text).

2.4 Principal component analysis

Principal component analysis (PCA) is a standard multivariate data analysis method (Esbensen and Swarbrick, 2018; Martens and Næs, 1989). It uses an orthogonal transformation to convert a set of observations often described by many correlated variables into a few linearly uncorrelated latent variables called principal components. We used the chemometric iterative NIPALS algorithm. Principal component analysis can alternatively be carried out based on singular value decomposition (SVD). The advantage of using SVD is that the code is simple, the disadvantage is that all possible components are always calculated, i.e., if X is large (i,e, is made up of a high number of variables), SVD is time consuming. NIPALS is slightly more complex to program, but a major advantage is that it allows to define the number of components to calculate. Thus, it saves time if X has many samples and variables. NIPALS also works on matrices with missing values; in that case SVD will fail. NIPALS is the standard chemometric approach and has been described in full detail in a wealth of literature (Esbensen and Swarbrick, 2018; Martens and Næs, 1989).

A series of spectral fingerprints (observations) can be assembled into an X-matrix with the 2048 FFT spectral frequencies as variables.

3   Results & discussion

Below follows the relevant PCA score-plots delineating the degree of discriminability obtained from running experiments pertaining to various factors and levels. In all the score plots, the average acoustic spectrum is shown to have a common reference point for all the score plots. For simplicity, only the results for paint 1 are shown for the investigations on the effect of the paint temperature, pressure and nozzle state (new, worn, clogged). The other paint types show similar results.

3.1 The effect of paint temperature

Figure 5 shows the score plot t1-t2 of the acoustic measurement at different temperatures. The baseline is indicated as "Average" while the low and high temperatures applied in the different groups of samples are labelled "Low Temp" and "High Temp", respectively. The groups are clearly separated, showing that the nozzle vibrations depend on the temperature. Both groups of samples lie to the right side of the base group, which suggests that the amplitude increases due to the flow through the nozzle. PC2 holds the information of the two sample groups with different temperature as their separation is mainly in the PC2 direction.
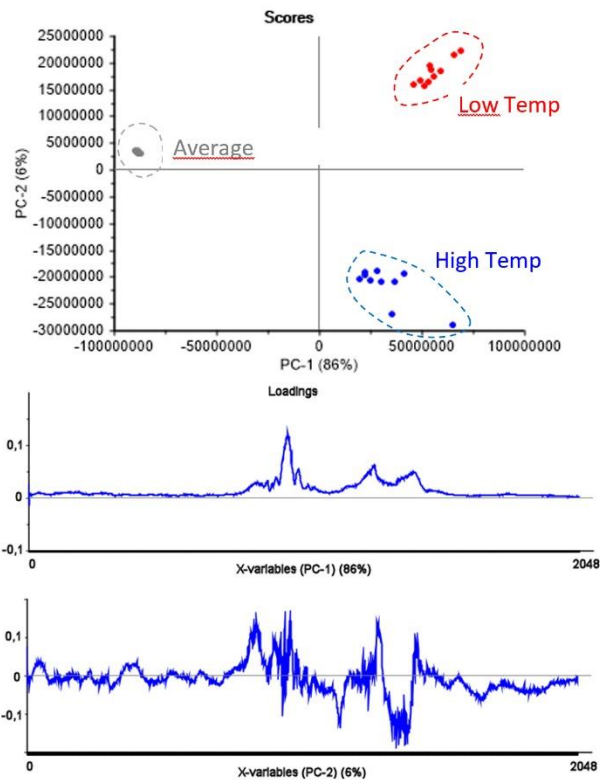
Figure 5. The effect of a temperature change, from top to bottom: Scores t1-t2, Loading 1, Loading 2

3.2 The effect of pump pressure

Figure 6 shows the score plot t1-t2 of the acoustic measurement at two different spraying pressures (high/low). Like for the effect of paint temperature, the baseline is also here indicated as "Average" while the low and high pressures are labelled "Low Pressure" and "High Pressure", respectively.

The groups are clearly separated, which indicates that the nozzle vibration depends on the paint pressure. Like in the score plot for the paint temperature, both groups of samples lie to the right side of the base group, indicating an increase in the amplitude due to the flow through the nozzle. PC2 holds the information of the two sample groups with different pressure as their separation is mainly in the PC2 direction.
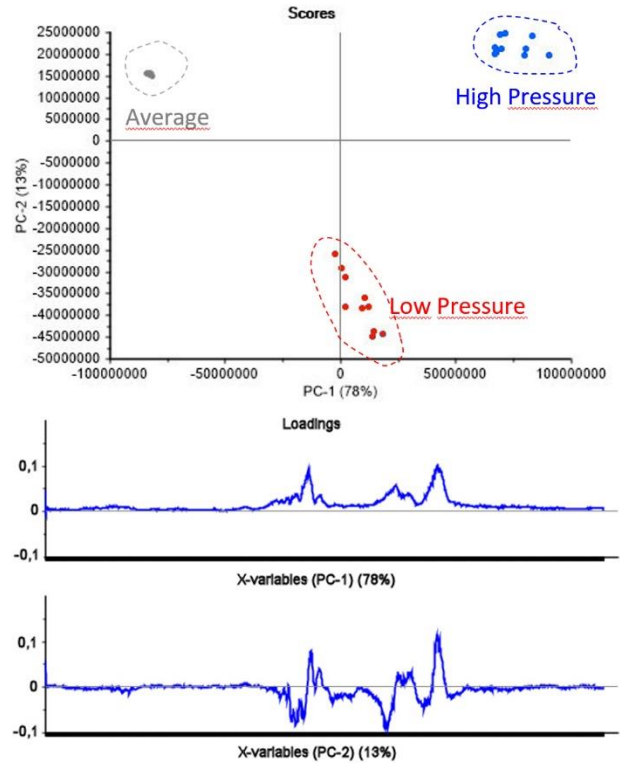
Figure 6. The effect of a change in pressure, from top to bottom: Scores t1-t2, Loading 1, Loading 2

3.3 The effect of nozzle wear

It is of high interest to monitor the state of a spraying nozzle as a worn nozzle changes the spray pattern. This can increase the loss of paint due to overspray and/or cause spraying defects. A new nozzle will typically last for 800 to 1500 liters of paint depending on spraying conditions and the paint type.

In Figure 7, two different nozzle conditions (new and worn) are shown in the score plot t1-t2 of the acoustic measurements. Like in Figure 5 and 6, the baseline is indicated as "Average" while the condition of the nozzle is indicated "new" and "worn".

The different location of the groups in the score plot indicates that the nozzle vibration depends on nozzle wear. The results are promising for the possibility of online monitoring of the nozzle state. A new nozzle will start with a negative score in the PC2 direction and slowly move to the direction of a positive score.
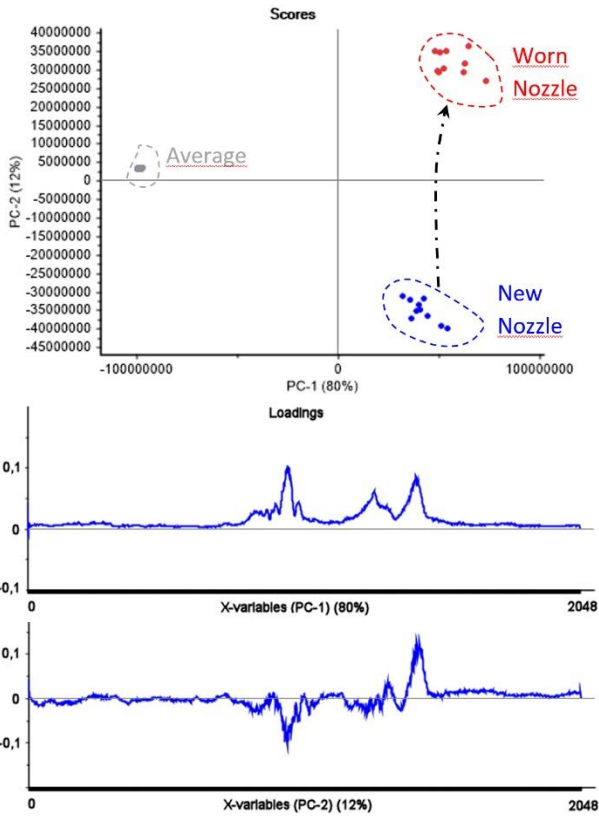
Figure 7. The effect of a nozzle wearing out, from top to bottom: Scores t1-t2, Loading 1, Loading 2
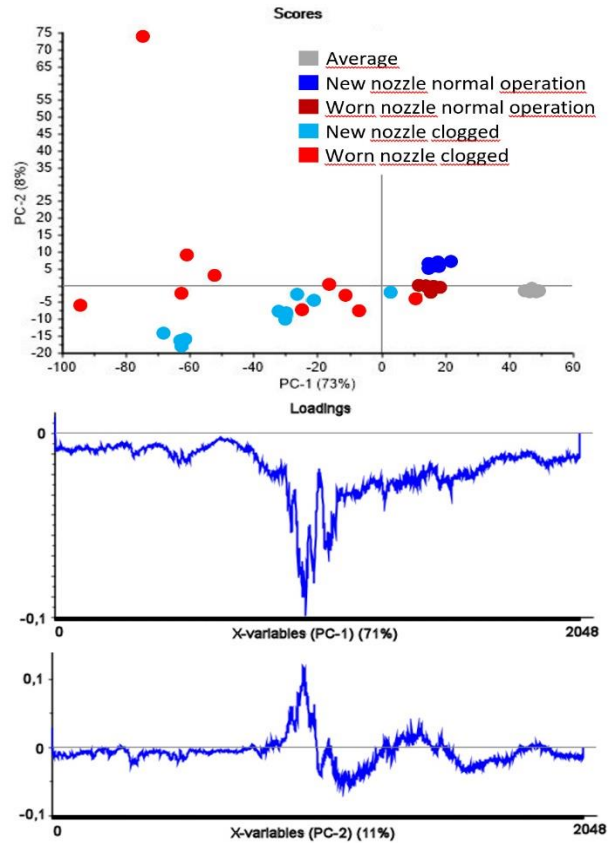
Figure 8. The effect of a clogged paint nozzle, from top to bottom Scores t1-t2, Loading 1, Loading 2

3.4 The effect of clogging

Another important aspect of process monitoring is the immediate detection of nozzle clogging because it causes coating defects and thus production loss. Figure 8 shows the results of a new and a worn nozzle in clogged and unclogged condition. Like in Fig. 7, the scores of a new and a worn nozzle in normal operation are clearly distinguishable from each other. The difference between scores of a clogged and an unclogged nozzle is even larger and allows an easy detection.

From these PCA results, there is a clear indication that the acoustic chemometric approach present the critical information needed to monitor the quality of paint spray evenness and nozzle condition during wood cladding.

3.5 The effect of paint type

Figure 9 shows the score plot t1-t2 of all four paint types, all the other variables such as temperature and pressure were kept constant. The score plot results suggest a clear influence of the paint type on the nozzle vibration. Paint 1 has the highest score along PC1, which indicates that this paint produces the highest sound amplitude. This can be probably ascribed to the high solid content and specific gravity of paint type 1 of 46 vol% and 1.45, respectively (Table 1).
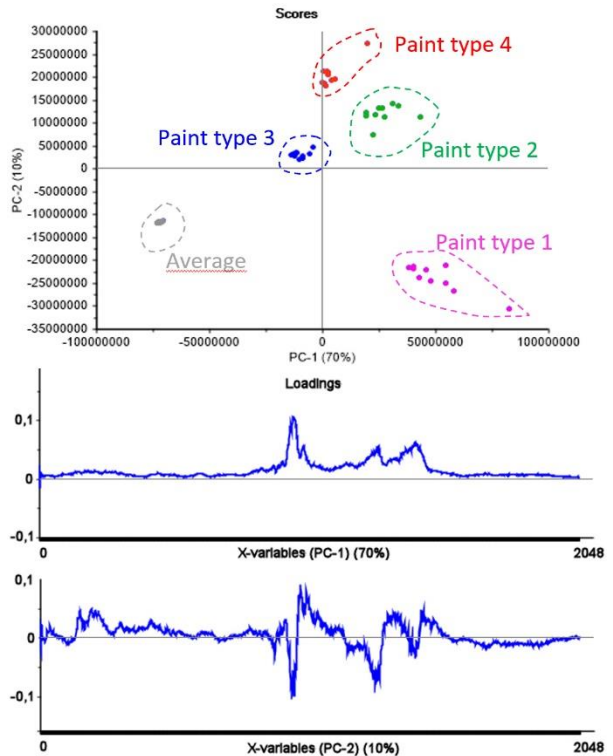

Figure 9. The effect of changing paint type, from top to bottom Scores t1-t2, Loading 1, Loading 2

4 Conclusion

The experimental setup based on accelerometer sensors with appropriate signal conditioning a.o. allows effective visualization of the different process states for different paints and working conditions using PCA score plots.

The present feasibility study shows the potential of our acoustic chemometric approach for the development of an industrial PAT paint spray monitoring system for wood coatings. However, there is a significant amount of on-site industrial calibration R&D necessary before a final method validation can be undertaken. The next steps will be to assess the technology in the harsh industrial environment of a coating plant for wood cladding. The present results rely on permanently affixed PAT sensors. Further studies will a.o. focus on the degree to which replacement of acoustic accelerometer sensors necessitate re-calibration of the multivariate data models employed – a critical issue for industrial implementation of acoustic chemometrics.

Finally, the applicability of acoustic chemometrics to other coating types and finishing processes than cladding, such as furniture, windows or flooring should be assessed.

Acknowledgements

References

B. K. Arvoh, S. Asdahl, K. Rabe, R. Ergon, and M. Halstensen. Online estimation of reject gas and liquid flow rates in compact flotation units for produced water treatment. *Flow Meas. Instrum.*, 24: 63–70, 2012. doi:10.1016/j.flowmeasinst.2012.03.008

K. A. Bakeev, *Process Analytical Technology, 2nd edition*. *Wiley*, Chichester, United Kingdom, 2010. ISBN: 978-0-470-72207-7, 2010, doi:10.1002/9780470689592

K. H. Esbensen, and B. Swarbrick, *Multivariate Data Analysis – An introduction to Multivariate Analysis, Process Analytical Technology and Quality by design*, 6th Ed., *CAMO Publishing*, 2018, ISBN 078-82-691104-0-1

K. H. Esbensen, B. Hope, T. T. Lied, M. Halstensen, T. Gravermoen, and K. Sundberg. Acoustic chemometrics for fluid flow quantifications: II. A small constriction will go a long way. *Journal of Chemometrics* 13 (27): 209–236, 1999. doi: 10.1002/ (SICI) 1099-128X (199905/08)13:3/4b209::AID-CEM553>3.0.CO;2–5.

K. H. Esbensen, M. Halstensen, T. T. Lied, A. Saudland, J. Svalestuen, S. de Silva, and B. Hope. Acoustic chemometrics — from noise to information. *Chemometr.*

Intell. Lab Syst. 44 (1–2): 61–76, 1998. doi:10.1016/S0169-7439 (98)00114-2.

G. Grüll, F. Tscherne, I. Spitaler, and B. Forsthuber, Comparison of wood coating durability in natural weathering and artificial weathering using fluorescent UV-lamps and water. *European Journal of Wood and Wood Products* 72(3): 367-376, 2014. doi:10.1007/s00107-014-0791-y

M. Halstensen, *Acoustic Chemometrics, Experimental multivariate sensor technology and development of system prototypes for industrial multi-phase characterisation: selected forays*. Ph.D. thesis, Telemark University College, Porsgrunn, Norway, 2001, ISBN 82-471-5381-5, ISSN 0809-103X

M. Halstensen, J. Lundberg, P.I. Januschas, and H.-P. Halvorsen, On-line Monitoring of Viscous Properties of Anti-icing Fluid Based on Partial Least Squares Regression Modeling, *Linköping Electronic Conference Proceedings*, 170:26-31, 2019. 10.3384/ecp2017026

M. Halstensen and K. H. Esbensen. New developments in acoustic chemometrics prediction of particle size distribution—the problem is the solution. *Journal of Chemometrics*, 14: 463–481, 2000. doi:10.1002/1099-128X(200009/12)14:5/6b463:: AID-CEM628>3.0.CO;2-Y.

M. Halstensen and K.H. Esbensen. Acoustic chemometric monitoring of industrial production processes in: K.A. Bakeev, *Process Analytical Technology,* Wiley, Chichester, United Kingdom, 2: 281–302, 2010. doi:10.1002/9780470689592, ISBN: 978-0-470-72207-7

M. Halstensen, P. de Bakker, and K. H. Esbensen. Acoustic chemometric monitoring of an industrial granulation production process—a PAT feasibility study. *Chemometr. Intell. Lab Syst,* 84: 88–97, 2006. doi:10.1016/j.chemolab.2006.05.012

M. Halstensen, S. R. de Silva, and K.H. Esbensen. Acoustic monitoring of pneumatic transport lines—from noise to information. *KONA Powder and Particle* 16 (8): 170–178 1998. doi:10.14356/KONA.1998020

U. Hundhausen, D. Kraniotis, S. Charisi, and M-L. Sortland, Quality control in coating lines for exterior cladding. In: *Proceedings International Woodcoatings Congress 2016.* 25-26, 2016. ISBN: 9783866302815

U. Hundhausen, M. Slabohm, and P. Meinlschmidt, Industrial coating of wood cladding: Inline control of board temperature, film thickness, and microfoam. In: *Proceedings International Woodcoatings Congress 2018.* 20-21, 2018. ISBN: 9783748600039

Emanuel Ifeachor and Barrie Jervis, *Digital Signal Processing, a Practical Approach*, *Addison-Wesley Publishing,* 1993. ISBN-13: 978-0201596199

F. N. Ihunegbo, M. Madsen, K. H. Esbensen, J. B. Holm-Nielsen, and M. Halstensen. Acoustic chemometric prediction of total solids in bioslurry: A full-scale feasibility study for on-line biogas process monitoring. *Chemometr. Intell. Lab. Syst.,* 110: 135–143, 2012. doi:10.1016/j.chemolab.2011.10.009

KonTre (Innovation Project for the Industrial Sector– BIONÆR, project no. 269206)

H. Martens and T. Næs. *Multivariate Calibration*, *Wiley*, Chichester, UK, 1989 reprint 1994. ISBN-10: 0471930474, ISBN-13: 978-0471930471

C. Wagner, F. N. Ihunegbo, M. Halstensen, and K.H. Esbensen. Acoustic chemometrics for material composition quantification in pneumatic conveying - The critical role of representative reference sampling. *Powder Technology*, 237: 506-513, 2013. doi:10.1016/j.powtec.2012.12.036

Notes on the Floquet-Lyapunov Theory of Linear Differential Systems

Kai Zenger

Department of Electrical Engineering and Automation, Aalto University, Finland, `kai.zenger@aalto.fi`

Abstract

This short remark describes some of the key points of the Floquet-Lyapunov theory of linear periodic differential systems. The basic transformation is presented, and its properties from the point of view of the source and target matrices and stability are discussed. The proposed methodology can be used in analysis, control design and simulation of systems with time-periodic characteristics. A method for designing a stabilizing state feedback control law is proposed. The methodology can be used in analysis and controller design for example in processes involving rotating machines and engines.

Keywords: time periodic systems, Floquet-Lyapunov transformation, monodromy matrix, stability

1 Introduction

The analysis and controller design of time-varying systems is difficult even in the case that the system or its model can be approximated to be linear. Such systems are for example rotating machines with changing speed or chemical reactors with changing flow rates and liquid volumes. Mathematically the models of such systems contain time-varying parameters. That is the case with the input-output mappings, which for continuous time differential models means differential equations or state-space representations.

The fundamental reason for the difficulties in analysis is that the solution of the time-varying differential system can usually not be expressed by elementary functions (Rugh, 1993). In system-theoretic terms this fact can be noted by the fact that the *state transition matrix* cannot be solved analytically. This function (or the solution of the system equation) exists but unlike in the time invariant case it cannot be expressed by elementary functions.

Input-output systems are usually modelled by state-space representations. Generally, the matrices involved (system matrix, input and output matrices) contain values that change with time. It is well-known that with a suitable transformation of the state variable (change of coordinates) exists such that the system matrix in the new coordinates becomes constant (Harris and Miles, 1980). Generally, the time-varying transformation matrix between the state variables is then complex, which is unfortunate when considering the potential use of the transformation for controller design. Anyway the original and the new

state-space representation are *kinematically similar* under the state transformation $x(t) = P(t)s(t)$, where $P(\cdot)$ is the transformation matrix (Harris and Miles, 1980). Further, if $P(\cdot)$ is a *Lyapunov transformation* the stability properties of the original and new state-space representation are known to be the same. The matrix is a Lyapunov transformation if it is invertible and both the norms of the matrix and its inverse are bounded for all time instants (Lyapunov, 1966). A fundamental problem is the fact that a suitable transformation is hard to find.

In short: Given a system matrix $A(\cdot)$ and choosing another matrix $E(\cdot)$ with the same dimension there exists a time-varying similarity transformation $P(\cdot)$ that changes $A(\cdot)$ to $E(\cdot)$. However, $P(\cdot)$ is not necessarily a Lyapunov transformation. Choosing $P(\cdot)$ in advance to be a Lyapunov transformation gives $E(\cdot)$ with the same stability properties as $A(\cdot)$. But $E(\cdot)$ is then determined by the given transformation and cannot be chosen freely,

The situation is slightly less complicated, if the original system matrix is *periodic*. Then it is known (Wiberg, 1971; Harris and Miles, 1980; Zenger and Ylinen, 2008) that a Lyapunov function exists which changes the original system matrix into a constant matrix. However, this function is not nesearily real. The purpose of this paper is to study real Lyapunov transformations, using homogenous systems first. Later, a synthesis approach is derived for designing stabilizing control algorithms for input-output systems. The general theory is called *Floquet-Lyapunov theory* due to G. Floquet who invented it and A. M. Lyapunov who investigated the stability of periodic systems. The basic ideas are presented first, leading to the very important concept of the *monodromy matrix*. The characteristics of the state transformation and the target (constant) system matrix are studied in detail, and it is shown how a given periodic Lyapunov function and constant system matrix E can be used to characterize the original time-varying system matrix and the solution of the related differential equation. The theory leads in a natural way to methods, where equations needed in synthesis problems, i.e. controller design problems, can be formulated.

2 Linear periodic differential systems

Consider the autonomous differential system

$$\dot{x}(t) = A(t)x(t), \qquad x(t_0) = x_0 \qquad (1)$$

where $x(\cdot) \in (\Re^n)^\Re$ and for all t it holds $A(t+T) = A(t)$ for some positive value T. The system matrix $A(\cdot) \in (\Re^{n \times n})^\Re$ is then T-periodic. (To be precise, T must be the smallest positive number that fulfils $A(t+T) = A(t)$.) It is assumed that the function $A(\cdot)$ is continuously differentiable, which guarantees that the solution to (1) is continuously differentiable as well. Actually, the assumption could be relaxed to allow piecewise continuous function $A(\cdot)$, meaning that the solution is still unique and continuous.

Let

$$x(t) = P(t)s(t) \qquad (2)$$

where $P(\cdot) \in (\Re^{n \times n})^\Re$ is a continuously differentiable and invertible function for all t. Assume also that both P and its inverse are bounded for all t. Easy calculations give (see (Bittanti, 2008; Zenger and Ylinen, 2008))

$$\dot{s}(t) = E(t)s(t) \qquad (3)$$

$$\dot{P}(t) = A(t)P(t) - P(t)E(t) \qquad (4)$$

$$P(t) = \Phi_A(t,t_0)P(t_0)\Phi_E(t,t_0)^{-1} = \Phi_A(t,0)P(0)\Phi_E(t,0)^{-1} \qquad (5)$$

$$\Phi_A(t,\tau) = P(t)\Phi_E(t,\tau)P(\tau)^{-1} \qquad (6)$$

where $\Phi_A(\cdot,\cdot)$ and $\Phi_E(\cdot,\cdot)$ are the state transition matrices related to $A(\cdot)$ (source matrix) and $E(\cdot)$ (target matrix), respectively. Note that the matrix $E(\cdot)$ is determined by $A(\cdot)$ and the transformation $P(\cdot)$. But it is also possible to choose the target system matrix $E(\cdot)$ freely (same dimension as with $A(\cdot)$), which then determines $P(\cdot)$.

The following question now arises: when is the matrix P T-periodic? To answer, note that

$$\begin{aligned} P(t+T) &= \Phi_A(t+T,t_0)P(t_0)\Phi_E(t+T,t_0)^{-1} \\ &= \Phi_A(t+T,t)\Phi_A(t,t_0)P(t_0)\left[\Phi_E(t+T,t)\Phi_E(t,t_0)\right]^{-1} \\ &= \Phi_A(t+T,t)\underbrace{\Phi_A(t,t_0)P(t_0)\Phi_E(t,t_0)^{-1}}_{P(t)}\Phi_E(t+T,t)^{-1} \end{aligned}$$

By introducing the *monodromy matrix*

$$\Psi_A(t) = \Phi_A(t+T,t) \qquad (7)$$

we obtain

$$P(t+T) = \Psi_A(t)P(t)\Psi_E(t)^{-1} \qquad (8)$$

The condition that $P(\cdot)$ is T-periodic is that for all t it holds $P(t+T) = P(t)$. Then

$$\Psi_A(t)P(t)\Psi_E(t)^{-1} = P(t) \Rightarrow \Psi_A(t) = P(t)\Psi_E(t)P(t)^{-1} \qquad (9)$$

or

$$\Phi_A(t+T,t) = P(t)\Phi_E(t+T,t)P(t)^{-1} \qquad (10)$$

which must hold for all t. The equation (10) is the condition that must be fulfilled for the transformation matrix $P(\cdot)$ to be T-periodic. Further, we can see from (9) that the monodromy matrices $\Psi_A(\cdot)$ and $\Psi_E(\cdot)$ are similar at

all time instants. That implies that their eigenvalues are the same at each time instant. Note that in the derivations of equations (9) and (10) no assumptions on the periodicity of the matrices $A(\cdot)$ and $E(\cdot)$ have been made. The result is then general.

But assuming that $A(\cdot)$ is T-periodic we can go further by noticing that

$$\begin{aligned} \Psi_A(t_1) &= \Phi_A(t_1+T,t_1) = \Phi_A(t_1+T,t_2)\Phi_A(t_2,t_1) \\ &= \Phi_A(t_1+T,t_2+t)\Phi_A(t_2+T,t_2)\Phi_A(t_2,t_1) \\ &= \Phi_A(t_1,t_2)\underbrace{\Phi_A(t_2+T,t_2)}_{\Psi_A(t_2)}\Phi_A(t_1,t_2)^{-1} \end{aligned}$$
$$(11)$$

where the important result that for the state transition matrix corresponding a T-periodic matrix $A(\cdot)$ it holds $\Phi_A(t_1+T,t_2+T) = \Phi_A(t_1,t_2)$ for all t_1,t_2 has been used. (See the Remark below.) It follows that at all time instants t_1 and t_2 the monodromy matrices $\Psi_A(t_1)$ and $\Psi_A(t_2)$ are similar. Hence, their eigenvalues are the same for all t meaning that the eigenvalues are constant.

So, if $P(\cdot)$ and $A(\cdot)$ are T-periodic, the eigenvalues of $\Psi_A(\cdot)$ and $\Psi_E(\cdot)$ are equal and constant.

Remark 1: The fact that the state transition matrix a T-periodic system matrix $A(t)$ fulfils $\Phi_A(t+T,\tau+T) = \Phi_A(t,\tau)$ for all t and τ follows by noticing that

$$\begin{aligned} \frac{\partial \Phi_A(t+T,\tau+T)}{\partial(t+T)} &= A(t+T)\Phi_A(t+T,\tau+T) \\ &= A(t)\Phi_A(t+T,\tau+T) \end{aligned}$$

$$\Phi_A(t+T,\tau+T)_{|t=\tau} = I$$

and on the other hand

$$\frac{\partial \Phi_A(t,\tau)}{\partial t} = A(t)\Phi_A(t,\tau)$$
$$\Phi_A(t,\tau)_{|t=\tau} = I$$

Both functions $\Phi_A(t+T,\tau+T)$, $\Phi_A(t,\tau)$ satisfy the same differential equation and have one common point (=I). Hence, the functions are equal.

Remark 2: Note that the condition for $P(\cdot)$ given in (5) holds generally in the sense that for all t, t_1

$$\begin{aligned} P(t) &= \Phi_A(t,t_1)\Phi_A(t_1,t_0)P(t_0)\left[\Phi_E(t,t_1)\Phi_E(t_1,t_0)\right]^{-1} \\ &= \Phi_A(t,t_1)\underbrace{\Phi_A(t_1,t_0)P(t_0)\Phi_E(t_1,t_0)^{-1}}_{P(t_1)}\Phi_E(t,t_1)^{-1} \\ &= \Phi_A(t,t_1)P(t_1)\Phi_E(t,t_1)^{-1} \end{aligned}$$

In short: provided that $A(\cdot)$ is T-periodic, $E(\cdot)$ can be chosen arbitrarily as long as it is T-periodic also. Then the condition that for all t (10)

$$\Phi_A(t+T,t) = P(t)\Phi_E(t+T,t)P(t)^{-1} \qquad (12)$$

guarantees the periodicity of $P(\cdot)$. Moreover, if the above condition is fulfilled at some time instant t_1, it is satisfied for all t.

If a constant target matrix E is chosen, the periodicity of $P(\cdot)$ is guaranteed by

$$\Phi_E(t+T,t) = e^{\int_t^{t+T} E\,d\tau} = e^{E(t+T-t)} = e^{ET} \qquad (13)$$
$$\Rightarrow e^{ET} = P(t)^{-1}\Phi_A(t+T,t)P(t)$$

which must hold for all t so without loss of generality it is possible to take $t = 0$

$$e^{ET} = P(0)^{-1}\Phi_A(T,0)P(0) \qquad (14)$$

Further, let us remember the *Jacobi-Liouville* formula

$$\det\Phi_A(t,\tau) = e^{\int_\tau^t \mathrm{tr}(A(v))\,dv} \qquad (15)$$

and use it in (12)

$$\det\Phi_A(t+T,t) = e^{\int_t^{t+T} \mathrm{tr}(A(v))\,dv} \qquad (16)$$
$$= \det P(t) \cdot \det\Phi_E(t+T,t) \cdot \underbrace{\det P(t)^{-1}}_{1/\det P(t)} \qquad (17)$$
$$= e^{\int_t^{t+T} \mathrm{tr}(E(v))\,dv} \qquad (18)$$

It follows that

$$\int_t^{t+T} \mathrm{tr}(A(v) - E(v))\,dv = 0 \qquad (19)$$

which must be fulfilled for all t. The result gives a necessary relationship between the periodic matrices $A(\cdot)$ and $E(\cdot)$. The condition is not sufficient.

It should be emphasized that although for a T-periodic system matrix $A(\cdot)$ it holds

$$\Phi_A(t+T,\tau+T) = \Phi_A(t,\tau) \qquad (20)$$

that does not mean that the solution $x(t)$ would be periodic. In fact

$$x(t+T) = \Phi_A(t+T,t_0)x_0 \qquad (21)$$
$$= \Phi_A(t+T,t_0+T)\Phi_A(t_0+T,t_0)x_0 \qquad (22)$$
$$= \Phi_A(t,t_0)\Psi_A(t_0)x_0 \qquad (23)$$

It follows that if one of the eigenvalues of the monodromy matrix $\Psi_A(t_0)$ is equal to 1, then choosing the initial condition x_0 to be the corresponding eigenvector leads to a periodic solution. Also, it is easy to see from the above equations that if a solution (state transition matrix) is known for one time period $[t_0, t_0 + T]$ it can then be determined for all t.

3 Stability of periodic systems

Stability of periodic systems is the natural question to be discussed next. To that end take E as a constant matrix and so

$$e^{ET} = P(t)^{-1}\Phi_A(t+T,t)P(t) = P(t)^{-1}\Psi_A(t)P(t) \quad (24)$$

or equivalently

$$e^{ET} = P(0)^{-1}\Phi_A(T,0)P(0) = P(0)^{-1}\Psi_A(0)P(0) \quad (25)$$

The condition guarantees that $P(\cdot)$ is periodic and thus a *Lyapunov transformation*, which is known to preserve the stability properties between the original system matrix $A(\cdot)$ and target $E(\cdot)$, (Rugh, 1993). The stability of E then determines the stability of $A(\cdot)$ as well. But the eigenvalues of $\exp(ET)$ (*characteristic multipliers*) are the constant eigenvalues of the monodromy matrix $\Psi_A(t)$. E is asymptotically stable, if and only if all its eigenvalues (*characteristic exponents*) are on the left half plane. But if λ_E is an eigenvalue of E and λ is the corresponding eigenvalue of $\exp(ET)$ then it holds

$$\lambda = e^{\lambda_E T} \qquad (26)$$

and it follows that the system is asymptotically stable, if and only if all eigenvalues of the monodromy matrix $\Psi_A(t)$ are inside the unit circle.

The same result can easily be obtained by looking at the properties of the monodromy matrix. For all real numbers τ and integer values k it holds

$$\Psi_A(\tau + (k+1)T) = \Psi_A(\tau) \qquad (27)$$

To see this, write

$$\Psi_A(\tau + (k+1)T) = \Phi_A(\tau + (k+2)T, \tau + (k+1)T)$$
$$= \Phi_A(\tau + (k+1)T, \tau + kT)$$
$$= \ldots = \Phi_A(\tau + T, \tau) = \Psi_A(\tau)$$

The solution of the state equation at time $\tau + (k+1)T$ becomes

$$x(\tau + (k+1)T) = \Phi_A(\tau + (k+1)T, \tau + kT)x(\tau + kT)$$
$$= \Psi_A(\tau + kT)x(\tau + kT) = \Psi_A(\tau)x(\tau + kT)$$
$$= \Psi_A(\tau)^2 x(\tau + (k-1)T)$$
$$= \ldots = \Psi_A(\tau)^{k+1}x(\tau)$$

The eigenvalues of the monodromy matrix must then clearly be less than one in magnitude, if the system is asymptotically stable.

Example: Consider the system matrix

$$A(t) = \begin{bmatrix} 1 & \cos(t) \\ 0 & 0 \end{bmatrix}$$

where clearly $T = 2\pi$. The state transition matrix is

$$\Phi_A(t,\tau)$$
$$= \begin{bmatrix} e^{t-\tau} & (1/2)e^{t-\tau}(\cos(\tau)-\sin(\tau))+(1/2)(\sin(t)-\cos(t)) \\ 0 & 1 \end{bmatrix}$$

The monodromy matrix is

$$\Psi_A(t) = \Phi_A(t+T,t)$$
$$= \begin{bmatrix} e^T & (1/2)e^T(\cos(t)-\sin(t))+f_1(t) \\ 0 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} e^T & (1/2)e^T(\cos(t)-\sin(t))+f_2(t) \\ 0 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} e^T & (1-e^T)f_2(t) \\ 0 & 1 \end{bmatrix}$$

where

$$f_1(t) = (1/2)(\sin(t+T)-\cos(t+T))$$

$$f_2(t) = (1/2)(\sin(t)-\cos(t))$$

The eigenvalues are e^T and 1, and the system is unstable. That can also be seen from the state transition matrix directly.

4 Control design

The above theory provides the background information and tools for analysing and simulating time-periodic differential systems. It was shown that any system matrices $A(\cdot)$ and $E(\cdot)$ with equal dimensions are equivalent to each other through a real transformation matrix $P(\cdot)$. However, it is very difficult to determine this matrix. Even worse, it is not clear that a such matrix would be stability preserving (Lyapunov transformation) between the source and target system matrices.

For periodic systems things turn out to be simpler. For example, the target system matrix can be a constant real matrix and the transformation a periodic Lyapunov transformation. According to equation (24) or

$$\Psi_A(t) = P(t)e^{ET}P^{-1}(t) \qquad (28)$$

it is possible to pick any real (stable or unstable) matrix E and a periodic Lyapunov transformation $P(\cdot)$ to calculate the monodromy matrix of the source system. This fact can be used to generate a stabilizing controller for time-periodic systems. To see that, consider the state-space representation of an input-output system

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) \quad x(t_0) = x_0$$
$$y(t) = C(t)x(t) + D(t)u(t) \qquad (29)$$

where the matrices $A(\cdot)$, $B(\cdot)$, $C(\cdot)$ and $D(\cdot)$ are time-periodic with the period T. Consider a time-varying state feedback control law

$$u(t) = -L(t)x(t)$$

where the gain matrix $L(\cdot)$ is T-periodic also. The closed-loop system can be written as

$$\dot{x}(t) = (A(t)-B(t)L(t))x(t) = A_{cl}(t)x(t)$$
$$y(t) = (C(t)-D(t)L(t))x(t) \qquad (30)$$

and the system matrix $A_{cl}(\cdot)$ is periodic.

The idea is to pick an asymptotically stable constant real matrix E and a Lyapunov transformation matrix $P(\cdot)$. If the controller gain is chosen such that the closed loop system matrix is kinematically similar to the matrix E, a stabilizing contol law has been found. The monodromy matrix of the closed loop (which does not even have to be calculated) can be represented as

$$\psi_{A_{cl}}(t) = P(t)e^{ET}P(t)^{-1} \qquad (31)$$

But now according to (4)

$$\dot{P}(t) = A_{cl}(t)P(t) - P(t)E$$
$$\Rightarrow A_{cl}(t) = \left(\dot{P}(t)+P(t)E\right)P(t)^{-1} \qquad (32)$$

and, using the expression for A_{cl}

$$B(t)L(t) = A(t) - A_{cl}(t)$$
$$= A(t) - \left(\dot{P}(t)+P(t)E\right)P(t)^{-1} \qquad (33)$$

the condition for $L(\cdot)$ has been formed. However, it should be noticed that the number of states is usually larger that the number of inputs, meaning that the equation does not necessarily have a solution. There are two alternatives: An approximative solution using pseudoinverse can be used or the matrices E and $P(\cdot)$ can be used as a parameterization for an existing solution for $L(\cdot)$. The former is a practical solution, which does not guarantee stability. The latter is not a direct algorithm for solving the controller gain, but shows the conditions (parameterization) which can be used. As such, it can be compared to the corresponding LMI conditions presented e.g. by (Bittanti, 2008).

5 Conclusion

A short introduction to the Floquet-Lyapunov theory of periodic linear differential systems was given in the paper. It was shown that the results are reasonably easy to derive by starting from a transformation between source and target differential systems. Although the theory is not new, see e.g. (Bittanti, 2008) it is believed to be more easy to come up with by considering the presented approach. Also, using the characteristics of the theory a way to design stabilizing control laws for periodic systems was presented. In a way, the method is used to parameterize system matrices of periodic differential systems. The results are easy to extend to periodic discrete-time systems as well.

References

Sergio Bittanti. *Periodic systems : filtering and control.* Springer, New York, 2008.

C. J. Harris and J. F. Miles. *Stability of Linear Systems: Some Aspects of Kinematic Similarity*. Academic Press, New York, 1980.

A. M. Lyapunov. *Stability of Motion*. Academic Press, New York, 1966.

W. J. Rugh. *Linear system theory*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 356 p., 1993. ISBN 0-13-555038-6.

D. M. Wiberg. *State Space and Linear Systems*. McGraw-Hill, New York, 1971.

K. Zenger and R. Ylinen. Representations with constant system matrices of linear time-periodic dynamical systems. In *17th World Congress of the International Federation of Automatic Control, Seoul, Korea*, 2008.

Adaptation framework for an industrial digital twin

Antti Koistinen   Markku Ohenoja   Jani Tomperi   Mika Ruusunen

Control Engineering, Environmental and Chemical Engineering Research Unit, University of Oulu, Finland,
{antti.koistinen,markku.ohenoja,jani.tomperi,mika.ruusunen}@oulu.fi

Abstract

Digital twins for performance-oriented applications in industrial environments require systematic model maintenance. Model adaptation requires efficient optimization tools and continuous evaluation of measurement quality. The adaptation and model performance evaluation are based on the modeling error, making the adaptation prone also to the measurement errors. In this paper, a framework for combining model adaptation and measurement quality assurance are discussed. Two examples with simulated industrial-scale biopharmaceutical penicillin fermentation are presented to illustrate the usability of the framework.

Keywords: digital twin, adaptation, framework, differential evolution

1 Introduction

Simulation tools have efficiently been applied in various engineering problems such as in process design and production planning. Currently, the simulation tools are also developed to real-time utilization. In process industries, the simulation models are, for example, used as an open-loop decision support tool (scenario simulation, prediction). These models can be considered as digital twins as essentially, "the digital twin is a virtual and computerized counterpart of a physical system that can be used to simulate it for various purposes, exploiting a real-time synchronization of the sensed data coming from the physical system". (Negri et al., 2017)

The digital twins can be classified in terms of their utilization in different tasks and areas, for example, design twins, performance twins and product or production twins. An important classification can be made in terms of real-time integration between the twin and its physical counterpart. Digital twin needs to have a closed-loop, automatic integration to the real process. Otherwise the correct term is a digital model (without integration) or a digital shadow (an open-loop integration). (Kritzinger et al., 2018)

Foreseen possibilities applying digital twins in continuous processes have been discussed for example in (Sun et al., 2017). In case of digital twins and shadows in these kinds of performance-oriented applications, it is crucial that the simulation model represents the real system continuously. Therefore, the performance of the digital model needs to be evaluated and the model needs to be updated automatically to cope with unseen or unmodeled changes.

The continuous updating, namely model adaptation requires efficient data-analysis and optimization tools. In general, the adaptation can be based on several techniques (Kadlec et al., 2011). However, both the real-time requirements and the model complexity (interconnected measurements and parameters) can make the adaptation problem challenging. Several methods have been presented to match the physical process with the digital model (Ohenoja et al., 2018; Friman and Airikka, 2012; Pietilä et al., 2013; Schirrmacher et al., 2009), but with limited insight on the whole problem.

This paper elaborates the overall picture on development and maintenance of a performance digital twin in industrial processes. In real systems, the actual parameters are not known. Hence, the whole adaptation (model performance evaluation) can only be dependent on the modeling error, making the adaptation prone also to the measurement errors. It is assumed that the adaptation involves a multivariable optimization problem. The computational issues, such as the real-time requirements are out of the scope of this presentation. Figure 1 presents the discussed framework. In Sections 2—4, the different elements of the framework are treated more detailed. In Section 5, simulated examples are presented.

2 Measurements as a foundation for adaptation framework

Modern industrial automation system includes measurements with varying characteristics; There can be sparse or even discontinuous data from the online analyzers and laboratory measurements providing product quality measurements, continuous automation system data with even sampling interval, and data with very high sampling frequency that needs to be preprocessed before used for decision making and process control. In addition, soft sensors are often used to infer measurements from harsh environments or otherwise difficult locations, or they can be used parallel to physical measurement device for validation and backup purposes.
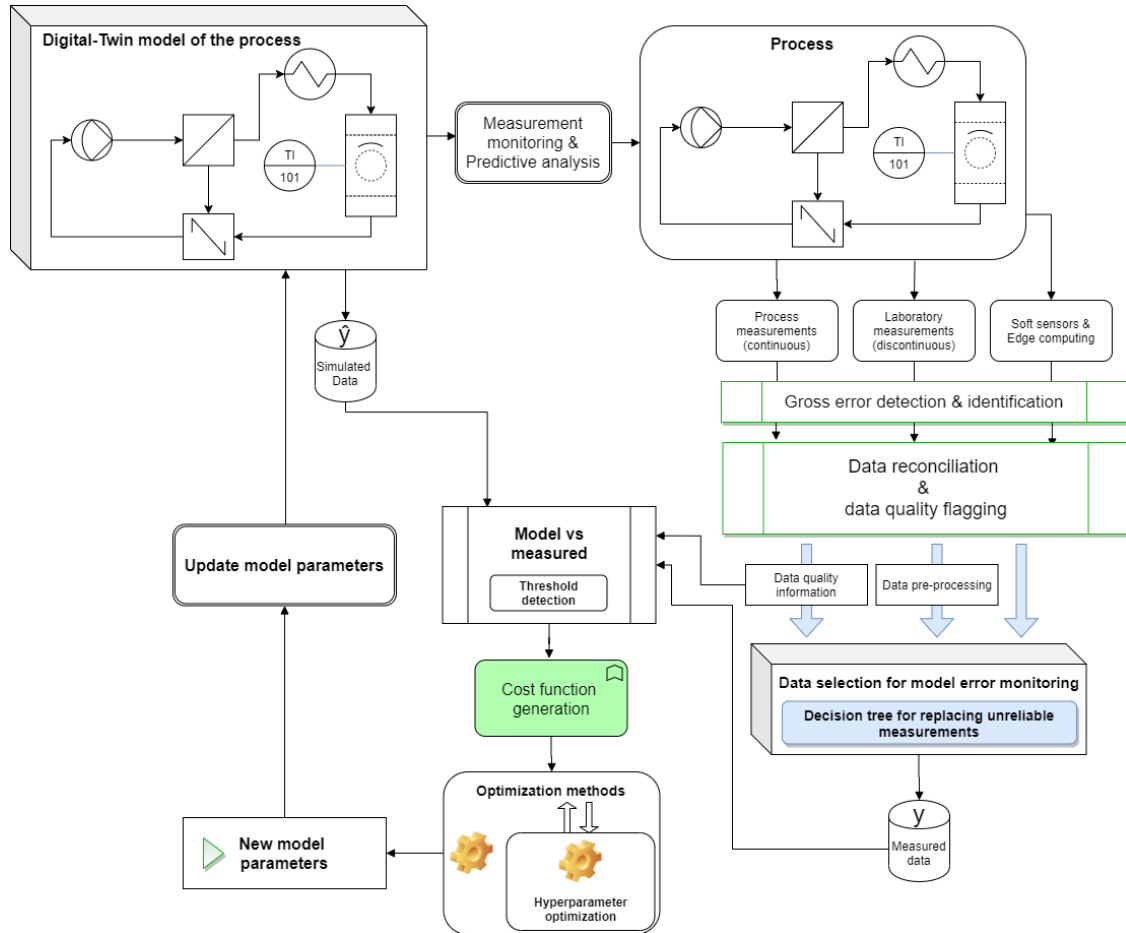
Figure 1. The concept for adaptation of industrial digital twin.

The concept for adaptation of industrial digital twin is presented in Figure 1. Different measurements from the real process on the right upper corner are used to monitor the modeling error of the digital twin on the upper left corner. Measurement quality is ensured by real-time monitoring and data reconciliation. Low quality or missing data is accounted in cost function generation.

Using a digital twin for decision making, robust self-diagnostics is required. Sensor malfunctions, measurement drifting, and other possible systematic error sources such as fouling may cause misinformation that leads to unreliable decisions. Measurement drift can be caused by ageing of components or environmental changes. Re-calibration is therefore important to maintain reliable information for a control system.

Soft sensors have always some error in their estimates or predictions. This error consists of modeling error and the quality of process measurements their model is based on. Similarly, physical sensors may perform acceptable at some defined measurement range, however their error can grow drastically when moving outside this working range.

Laboratory measurements are rarely used in automated control but can work as a basis for evaluating reliable process operation and for validating process measurements, as well building soft sensors. They are used as a basis for reliable decision making in many industrial processes. Sampling can induce large errors or variation in results if done improperly and inconsistently. Sample preparation and analysis should be done according to previously validated standard procedure to ensure the quality of data.

Intelligent measurement devices can detect if their data is reliable and possibly self-calibrate or alternatively alert the process operator or maintenance personnel. Modern automation system can include functionalities for detecting unreliable measurements and this is important aspect to consider when implementing digital twin. It is preferable to associate the information of the measurement quality with the measurement value for any kind of modeling and decision support.

3 Quality of measurements

Reliable measurement information is a prerequisite for the successful model adaptation. This requires online measurement quality monitoring. Unfit measurements can be removed or replaced. There are several methods for detecting differing measurement errors presented in chapter 3.1. If gross errors (Human mistakes, measurement errors, etc.) are also present in the process

data, they must be identified and removed before data reconciliation (Sánchez et al., 2002).

Intelligent measurement devices with vast amount of data and real-time data processing demands require automated quality assurance (QA) and quality control (QC) methods. (Campbell et al., 2013; Morello et al., 2014) A high performance control system includes automatic quality control algorithms and a comprehensive flagging system to indicate data quality level, to identify errors, to highlight corrected values, and to make it easier for data users to identify suspicious and erroneous data, namely to make quality control more efficient and closer to real-time. (Vejen et al., 2002)

3.1 Gross error detection & identification

Measurement drift is a typical for phenomenon related to industrial measurements where bias between true and measured values is evolving during long period of time. Baena-García et al. (2005) presented two methods for drift detection; Drift detection method uses binomial distribution to calculate the number of errors. On the other hand, early drift detection method considers the distance between two errors classification instead of considering only the number of errors.

Nishida and Yamauchi (2007) considers previous methods and presents new more noise resilient option which works for sudden concept shifts and gradual changes. They continued their work in (Nishida and Yamauchi, 2009) and presented learning system that uses multiple online classifiers that can predict changes. In Dries and Rückert (2009), six methods for online concept drift detection were evaluated and three approaches that can detect data drift reliably were presented. Finally, different gross error detection & identification (GEDI) methods, based e.g. in Mean squared error (MSE), root mean squared errors (RMSE), correlation coefficient based methods etc. are presented in (Kadlec et al., 2011).

3.2 Quality flagging

QA and QC procedures are closely related, but each has a distinct meaning. QA is a process of data profiling to discover inconsistencies and other anomalies in the data and performing data processing to improve the data quality. On the other hand, the QC process decides whether data meet the requirements for quality outlined by the end users. Hereby, QA can be considered a proactive or preventive process to avoid problems and QC as a process to identify and flag suspect data after they have been generated. (Campbell et al., 2013; Scully-Allison et al., 2018)

The QC procedures can be applied at various stages of data flow from sensors to the end user and can be carried out by numerous methods. Observations can be flagged by several methods and various symbols and names can be used to indicate the quality control level. It is difficult to develop common guidelines for QC and flagging procedures that are applicable in all circumstances and so there are no universal standards, but they are specific to the type of the data, application and the location at which data is collected. (Campbell et al., 2013; Vejen et al., 2002). Hence, several QC methods for automatic tests have been reported (Vejen et al. 2002; Scully-Allison et al. 2018; Geuder et al., 2015; Lewis et al., 2018)

In addition to the simple and traditional QC procedures, methods utilizing machine-learning have also been developed. These methods represent a data-driven approach to QC, wherein statistical models or classifiers are trained using empirical data collected from sensors. Hence, little knowledge is required about the device hardware or the phenomena being measured. On the other hand, adequate amount of data that contains the examples of faulty and correct data for model training and validation is required. Artificial neural networks, support vector machines, decision trees, and probabilistic models among others are commonly utilized as machine-learning approaches (Kadlec et al., 2011; Campbell et al., 2013; McNutt et al., 2019) but also hybrid systems, for example combining QC flags and a fuzzy logic, are developed (Morello et al., 2014).

Defective or missing data are unavoidable and require decisions how to process the faulty data: should the erroneous values to be removed, adjusted, replaced with an estimated value (e.g. soft sensor) or ignored. Great care must be taken to ensure that all processing steps are well documented so that they can be evaluated. The raw unmanipulated data should also always be saved. The uncertainties may arise from missing data and for instance using the wrong methods to fill the gaps. In a quality control procedure, all invalid data should be marked, but it is essential to ensure that valid data are not marked or removed, for example when a real but rare and extreme value outside the expected range occurs. In device self-diagnostics one difficulty that often arises is in differentiating between normal deviations and component faults.

Uncertainty techniques can be applied to measure the impact of faults on measurement quality, which makes it possible in certain circumstances to continue to use a sensor after it has developed a fault. Setting QA/QC tolerances to minimize false detections is difficult, especially under changing conditions. Moreover, the values of many parameters are site-dependent. In many cases, too many or too few events may be detected and therefore, the results of automatic screening demand a manual check of an expert to ensure the validity. Expert knowledge and analyzing the circumstances under which false errors occur provides information that can be used to adjust the QC procedure and achieve more optimal performance. (Campbell et al., 2013; Geuder et al., 2015)

3.3 Data reconciliation

One possibility to monitor the quality of measurements is based on data reconciliation. Data reconciliation also offers means to correct or replace the erroneous and missing measurements. (Vasebi et al., 2014)

The main idea of data reconciliation is to adjust the measurements data to satisfy the mass, energy, or momentum balance equations. Steady-state data reconciliation is solved as an optimization problem, where the objective is to minimize the difference between the measured variables and the adjusted ones, weighted by the reciprocal of the variance. Objective function can be e.g. weighted least squares function (WLS). (Vasebi et al., 2014)

4 Adaptation

In real systems, the actual model parameters are not known. Hence, the whole adaptation (model performance evaluation) can only be dependent on the modeling error, making the adaptation prone also to the measurement errors mentioned in the previous Section. Therefore, the adaptation should preferably utilize the uncertainty estimates and quality flags of the measurements. In addition, the adaptation framework needs to consider when the adaptation is needed, how to cope with varying data quality and how to make the trade-off between the adaptation accuracy, stability, and computational load.

4.1 Triggering

The adaptation algorithm can be run consistently or there can be predetermined threshold based on the estimation error. The regular adaptation interval can be determined based, for example, on the minimum process delay or maximum sampling interval of the relevant measurements.

The thresholds for infrequent adaptation can be calculated between the actual process measurements and the model output values representing the measurements. Methods for triggering should be chosen according to case. For example, (Palomo et al., 1991) mentions residual analysis, variance distribution and coherency-based methods among others. Also, in this case, the quality information of the measurements can be utilized to guide the triggering.

4.2 Optimization methods

There is a vast amount of optimization methods. On the other hand, it has been stated that there is a lack of a single solver that can overperformance the others in variety of optimization problems (Rios and Sahinidis, 2013). The performance of solvers is strongly dependent on the problem dimensionality and non-smoothness of the function and bounds on the variables.

In typical engineering problems, the global optimum is a solution that outperforms its alternatives after a fixed number of cost function evaluations. It has been shown that the metaheuristic methods can overperform the deterministic methods especially with small budgets (Sergeyev et al., 2018). On the other hand, the deterministic methods have provable convergence abilities to any optimization problems with an unlimited budget, but metaheuristic method may not be able to find the global optimum despite increased budget. (Sergeyev et al., 2018)

One well-known metaheuristic (stochastic) optimization method is the Differential Evolution (DE) (Storn and Price, 1997). Its implementation is sufficiently straightforward, and DE has only several hyperparameters to tune the algorithm. DE is utilized in the adaptation examples in this paper, also since calculations proceed with real numbers. The optimization method selection was done in the previous study (Ohenoja et al., 2018).

4.3 Selection of model parameters and outputs

In large-scale processes and simulation models, the number of adjustable model parameters, together with number of possible measurements (model outputs) gets high and the optimization problem becomes too complex. It is important to be able to focus to the most relevant parameters and measurements to decrease the problem dimensions. Naturally, expert knowledge is needed to perform the selection of parameters and measurements to the model adaptation.

Systematic approach to reach a smaller subset of possible parameters and measurements could involve sensitivity analysis tools. In local sensitivity analysis (LSA), one individual model parameter is changed at the time and the effect to model outputs observed. However, this approach cannot account for parameter interactions. Global sensitivity analysis (GSA) for estimating input parameter effects on different process outputs can use, for example, Sobol method (Sobol′, 2001). It is based on variability observed from Monte Carlo simulation and therefore large-scale system may require extensive computational cost.

4.4 Cost function selection

In model adaptation the objective is to match the model outputs with real measurements. Hence, an intuitive selection of the cost function is the sum of output measurement errors. However, the cost function typically contains measurement values with totally different ranges and the measurements that have largest error, will define the model parameters that are preferred/overweighed in the adaptation. Therefore, the first requirement is to normalize the values from the data sources to compensate the different absolute values.

Secondly, the adaptation performance can be improved by weighting certain measurements according to their importance to selected model input parameters. The weighting can be based, for example, on expert knowledge or systematic testing of model parameter sensitivities.

Finally, the adaptation framework for an industrial digital twin discussed in this paper indicates that the quality information connected to measurements should also be utilized in the cost function formulation. The cost function measurement weighting can be changed accordingly if quality flagging describes measurement as unreliable. Alternatively, this unreliable measurement can be replaced with soft sensor if available.

4.5 Hyperparameters

The selection of hyperparameters of the optimization method is crucial for its convergence and computational performance. Some rule of thumbs exists for different optimization methods, but typically the selection is based on experience and intensive off-line testing. In some scenarios, it would be beneficial to automatically change these algorithm tuning parameters to compensate decreased data quality or when the optimization performance drops noticeably. In the case of DE, it can be done using the method itself for this task. First loop optimizes the model parameters and the second loop finds the best model parameters for the task. (Chachuat et al., 2009)

5 Illustrative simulation examples

5.1 Simulation model

Industrial Penicillin Simulator (IndPenSim) is a MATLAB® based model for industrial-scale biopharmaceutical penicillin fermentation simulation. IndPenSim constitutes of several m-files providing their own function for the simulator. This program is intended to be a benchmark simulator for researchers to compare and validate new controllers in the batch fermentation process. It also includes Raman spectroscopy simulation that can be used for advanced process control design. It has fault generation capability built-in so it can be used for testing fault detection algorithms and for process fault identification. (Goldrick et al., 2019; 2015)

This paper uses a modified version of the simulator. Measurement noise and random variations were removed to move towards deterministic model. This serves our purpose of testing adaptation algorithm with adaptive cost function. The effect of random noise and measurement faults will be studied in the future after we understand how this algorithm performs in optimal conditions.

The IndPenSim includes 105 model parameters in total and 67 model outputs. The number of model outputs were decreased to 17 based on expert knowledge. The use of expert knowledge here refers to consideration of which of those model outputs could be measured directly or indirectly in real-time without discontinuous laboratory sampling. The model parameters were narrowed down to six by using LSA. The system of six inputs and 17 outputs was then

utilized in model adaptation and it was also executable for GSA.

5.2 Example with weighted objective function

This example aims to illustrate the effect of measurement weighting in the objective function (OF) of model adaptation. GSA was used to generate weighting factors for model outputs which represent different process measurements. Fractional sensitivity indices between the different model parameters and outputs were added up to form the weighting coefficient for each output. In the other scenario, all measurements were equally weighted.

The real process in this text refers to modeling results gained by a reference simulation with altered model parameters. Nominal model input parameters are used to represent the state of the digital twin to be adapted. The model input parameters are changed by 5% upwards to represent changed operating conditions of the real process. Adaptation algorithm is then used to find these new values using the error between process output measurements and model outputs.

Adaptation algorithm, realized with DE, compares the model output values from the simulated process to values received from different DE trials and starts to adapt the designated model parameters in an iterative way. Adaptation stops when maximum number of iterations is reached or if the cost function value reaches the predetermined threshold value (0.001), which relates to the total relative modelling error of 0.1%.

The cost function combines total relative difference between 17 model outputs and process measurements. The total error is a sum from a simulation period of 1150 samples. Adaptation performance is inspected using cumulative and dynamic output errors. In addition to model outputs, the adapted model parameter errors to known values are inspected (although not relevant in industrial implementations).

Preliminary work for this study includes selecting optimal adaptation window for minimizing the delay between adaptation cycles, while still maintaining a good performance. It was determined that 200 datapoints was enough to maintain good performance.

The results are presented in Figure 2 and Figure 3. Weighted OF lowers the total error in almost every measurement compared to nominal OF in Figure 2. Measurements with larger error compared to nominal OF have only small rise in total error while 'NH3', 'PAA', 'a3', 'a4', and 'S' have significant drops in total error. The total error with weighted OF is only 39% from the total error with nominal OF. The total error is still mainly influenced by single model output errors peaking at certain points of a batch process. Overall, all the measurements stay below the generally acceptable 5% error limit for the soft sensors, which indicates that the adaptation was successful.

According to results in Figure 3, the model parameters 'ratio_mu_e_mu_b' and 'K_diff' are the only ones to have larger error with weighted OF than with nominal OF. 'Y_PAA_P' has very large drop in error after measurement weighting.
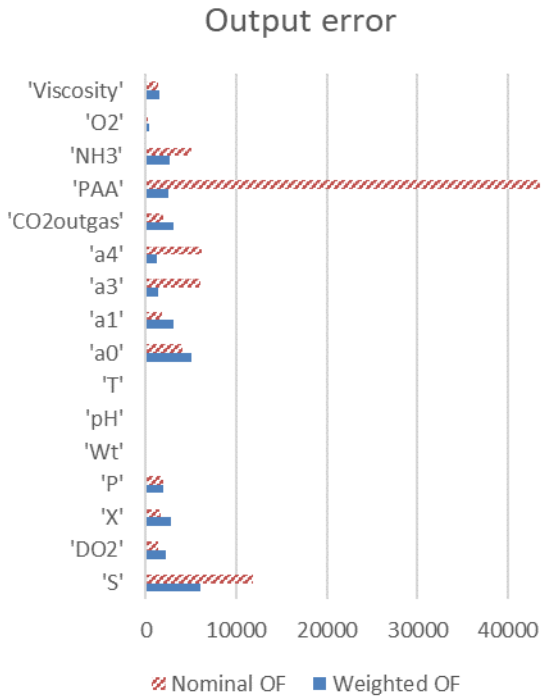
Figure 2. Total measurement errors between the model outputs and simulated process measurements using nominal and weighted output functions. The total error is a sum from a simulation period of 1150 samples.
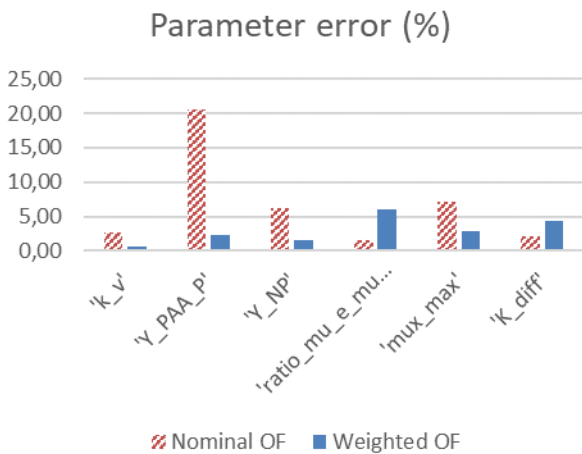
Figure 3. Model input parameter errors with nominal and weighted output functions.

5.3 Example with missing measurement

Adaptation performance is strongly dependent on the quality of measurement information. Loss of vital measurement information increases the adaptation error to a level where model is unreliable. This highlights the need for measurement quality monitoring and flagging. Adaptation robustness can be increased by implementing soft sensors that can be used to replace missing or faulty measurements.

Such a scenario is simulated here, where a vital measurement 'a0' is missing and replaced with a soft sensor. The soft sensor performs somewhat worse than the original measurement, as depicted in Figure 4.
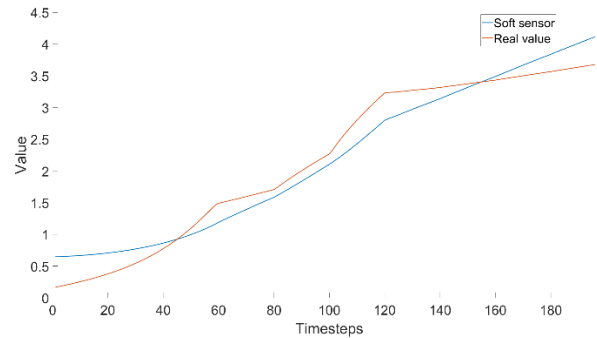
Figure 4. The comparison for soft sensor and real measurement in a0 value. The x-axis presents 200 samples with 12-minute sampling interval.

The adaptation result is presented in Figure 5. The missing measurement increases the output error tremendously. By replacing the missing measurement with a soft sensor, the adaptation performance can be maintained and the total errors in the model outputs can be noticeably decreases. Large errors in 'a3' and 'a4' are caused by the differences in measurement dynamics between the soft sensor and the real measurement.
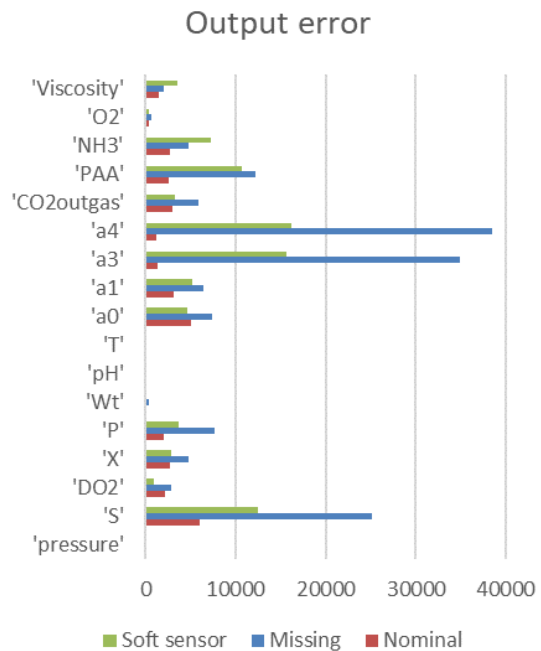
Figure 5. The effect of missing measurement information to adaptation performance.

6 Discussion and conclusions

The simulated results with modified IndPenSim are deterministic to some extent to be suitable for controlled testing of our adaptation framework. These results revealed the applicability of our LSA/GSA based approach for selecting the set of adapted model parameters and using GSA results for measurement weighting.

It was also depicted that the adaptation can be very dependent on a single measurement. This emphasize the importance of the reliable measurement information and real-time quality monitoring. Robustness can be improved by implementing data reconciliation and soft sensors to replace unreliable or missing measurements.

Work will continue with testing the adaptation with different measurement errors and by efficiently combining the measurement QA/QC information to model adaptation. The adaptation framework will also be stressed in mineral processing simulation models that represent real life process in Oulu Mining School enrichment pilot plant.

Acknowledgements

References

Manuel Baena-Garcıa, Ricard Gavalda, and Rafael Morales-Bueno. Early Drift Detection Method. *Computer Science*, 10, 2005.

John L. Campbell, Lindsey E. Rustad, John H. Porter, Jeffrey R. Taylor, Ethan W. Dereszynski, James B. Shanley, Corinna Gries, et al. Quantity Is Nothing without Quality: Automated QA/QC for Streaming Environmental Sensor Data. *BioScience* 63 (7): 574–85, 2013. doi:10.1525/bio.2013.63.7.10.

B. Chachuat, B. Srinivasan, and D. Bonvin. Adaptation Strategies for Real-Time Optimization. *Computers & Chemical Engineering*, Selected Papers from the 18th European Symposium on Computer Aided Process Engineering (ESCAPE-18), 33 (10): 1557–67, 2009. doi:10.1016/j.compchemeng.2009.04.014.

Anton Dries and Ulrich Rückert. Adaptive Concept Drift Detection. *Statistical Analysis and Data Mining: The ASA Data Science Journal* 2 (5–6): 311–27, 2009. doi:10.1002/sam.10054.

Mats Friman and Pasi Airikka. Tracking Simulation Based on PI Controllers and Autotuning. *IFAC Proceedings Volumes*, 2nd IFAC Conference on Advances in PID Control, 45 (3): 548–53, 2012. doi:10.3182/20120328-3-IT-3014.00093.

N. Geuder, F. Wolfertstetter, S. Wilbert, D. Schüler, R. Affolter, B. Kraas, E. Lüpfert, and B. Espinar. Screening and Flagging of Solar Irradiation and Ancillary Meteorological Data. *Energy Procedia*, International Conference on Concentrating Solar Power and Chemical Energy Systems, SolarPACES 2014, 69: 1989–98, 2015. doi:10.1016/j.egypro.2015.03.205.

Stephen Goldrick, Carlos A. Duran-Villalobos, Karolis Jankauskas, David Lovett, Suzanne S. Farid, and Barry Lennox. Modern Day Monitoring and Control Challenges Outlined on an Industrial-Scale Benchmark Fermentation Process. *Computers & Chemical Engineering*, 2019. doi:10.1016/j.compchemeng.2019.05.037.

Stephen Goldrick, Andrei Ştefan, David Lovett, Gary Montague, and Barry Lennox. The Development of an Industrial-Scale Fed-Batch Fermentation Simulation. *Journal of Biotechnology* 193: 70–82, 2015. doi:10.1016/j.jbiotec.2014.10.029.

Petr Kadlec, Ratko Grbić, and Bogdan Gabrys. Review of Adaptation Mechanisms for Data-Driven Soft Sensors. *Computers & Chemical Engineering* 35 (1): 1–24, 2011. doi:10.1016/j.compchemeng.2010.07.034.

Werner Kritzinger, Matthias Karner, Georg Traar, Jan Henjes, and Wilfried Sihn. Digital Twin in Manufacturing: A Categorical Literature Review and Classification. *IFAC-PapersOnLine*, 16th IFAC Symposium on Information Control Problems in Manufacturing INCOM 2018, 51 (11): 1016–22, 2018. doi:10.1016/j.ifacol.2018.08.474.

Elizabeth Lewis, Niall Quinn, Stephen Blenkinsop, Hayley J. Fowler, Jim Freer, Maliko Tanguy, Olivia Hitt, Gemma Coxon, Paul Bates, and Ross Woods. A Rule Based Quality Control Method for Hourly Rainfall Data and a 1 km Resolution Gridded Hourly Rainfall Dataset for Great Britain: CEH-GEAR1hr. *Journal of Hydrology* 564: 930–43, 2018. doi:10.1016/j.jhydrol.2018.07.034.

Todd R. McNutt, Kevin L. Moore, Binbin Wu, and Jean L. Wright. Use of Big Data for Quality Assurance in Radiation Therapy. *Seminars in Radiation Oncology*, Big Data in Radiation Oncology, 29 (4): 326–32, 2019. doi:10.1016/j.semradonc.2019.05.006.

Elisabetta B. Morello, Guillaume Galibert, Daniel Smith, Ken R. Ridgway, Ben Howell, Dirk Slawinski, Greg P. Timms, Karen Evans, and Timothy P. Lynch. Quality Control (QC) Procedures for Australia's National Reference Station's Sensor Data—Comparing Semi-Autonomous Systems to an Expert Oceanographer. *Methods in Oceanography* 9: 17–33, 2014. doi:10.1016/j.mio.2014.09.001.

Elisa Negri, Luca Fumagalli, and Marco Macchi. A Review of the Roles of Digital Twin in CPS-Based Production Systems. *Procedia Manufacturing*, 27th International Conference on Flexible Automation and Intelligent Manufacturing, FAIM2017, 27-30 June 2017, Modena, Italy, 11: 939–48, 2017. doi:10.1016/j.promfg.2017.07.198.

K. Nishida and K. Yamauchi. Learning, Detecting, Understanding, and Predicting Concept Changes. In *2009 International Joint Conference on Neural Networks*, 2280–87, 2009. doi:10.1109/IJCNN.2009.5178619.

Kyosuke Nishida and Koichiro Yamauchi. Detecting Concept Drift Using Statistical Testing. In *Discovery Science*, edited by Vincent Corruble, Masayuki Takeda, and Einoshin Suzuki, 264–69. Lecture Notes in Computer Science. Springer Berlin Heidelberg 2007.

Markku Ohenoja, Mika Ruusunen, Antti Koistinen, Jani Kaartinen, Janne Paaso, Ari Isokangas, and Marko Paavola. Towards Mineral Beneficiation Process Chain Intensification. *IFAC-PapersOnLine* 51 (21): 201–6, 2018. doi:10.1016/j.ifacol.2018.09.418.

E Palomo, J Marco, and H Madsem. METHODS TO COMPARE MEASUREMENTS AND SIMULATIONS. In , 570–77. Nice, France 1991.

Janne Pietilä, Jani Kaartinen, and Ari-Matti Reinsalo. Parameter Estimation for a Flotation Process Tracking Simulator. *IFAC Proceedings Volumes*, 16th IFAC Symposium on Control, Optimization and Automation in Mining, Minerals and Metal Processing, 46 (16): 122–27, 2013. doi:10.3182/20130825-4-US-2038.00048.

Luis Miguel Rios and Nikolaos V. Sahinidis. Derivative-Free Optimization: A Review of Algorithms and Comparison of Software Implementations. *Journal of Global Optimization* 56 (3): 1247–93, 2013. doi:10.1007/s10898-012-9951-y.

M. Sánchez, D. Leung, H. Konecsny, C. Bigaran, and J. A. Romagnoli. Smart Enterprise for Pulp & Paper Mills: Data Processing and Reconciliation. In *Computer Aided Chemical Engineering*, edited by Johan Grievink and Jan van Schijndel, 10:763–68. European Symposium on Computer Aided Process Engineering-12. Elsevier, 2002. doi:10.1016/S1570-7946(02)80155-9.

Vasco Schirrmacher, Philipp Schmiechen, Mirko Knaak, Akira Ohata, and Satoshi Kaneko. An Approach for Physical Model Adaptation Based on Transient Measurements. In *2009 ICCAS-SICE*, 480–83, 2009.

Connor Scully-Allison, Vinh Le, Eric Fritzinger, Scotty Strachan, Frederick C. Harris, and Sergiu M. Dascalu. Near Real-Time Autonomous Quality Control for Streaming Environmental Sensor Data. *Procedia Computer Science*, Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 22nd International Conference, KES-2018, Belgrade, Serbia, 126: 1656–65, 2018. doi:10.1016/j.procs.2018.08.139.

Ya D. Sergeyev, D. E. Kvasov, and M. S. Mukhametzhanov. On the Efficiency of Nature-Inspired Metaheuristics in Expensive Global Optimization with Limited Budget. *Scientific Reports* 8 (1): 453, 2018. doi:10.1038/s41598-017-18940-4.

I.M Sobol′. Global Sensitivity Indices for Nonlinear Mathematical Models and Their Monte Carlo Estimates. *Mathematics and Computers in Simulation* 55 (1–3): 271–80, 2001. doi:10.1016/S0378-4754(00)00270-6.

Rainer Storn and Kenneth Price. Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. *Journal of Global Optimization* 11 (4): 341–59, 1997. doi:10.1023/A:1008202821328.

Bei Sun, Sirkka-Liisa Jämsä-Jounela, Yancho Todorov, Laurentz E. Olivier, and Ian K. Craig. Perspective for Equipment Automation in Process Industries. *IFAC-PapersOnLine*, Control Conference Africa CCA 2017, 50 (2): 65–70, 2017. doi:10.1016/j.ifacol.2017.12.012.

Amir Vasebi, Éric Poulin, and Daniel Hodouin. Selecting Proper Uncertainty Model for Steady-State Data Reconciliation – Application to Mineral and Metal Processing Industries. *Minerals Engineering* 65: 130–44, 2014. doi:10.1016/j.mineng.2014.06.001.

Flemming Vejen, Caje Jacobsson, Ulf Fredriksson, Margareth Moe, Lars Andresen, Eino Hellsten, Pauli Rissanen, Þóranna Pálsdóttir, and Þordur Arason. Quality Control of Meteorological Observations. Automatic Methods Used in the Nordic Countries. CLIMATE REPORT 8/2002 KLIMA. Oslo: NORWEGIAN METEOROLOGICAL INSTITUTE 2002.

Sensor fault detection with Bayesian networks

Esin Iplik   Ioanna Aslanidou   Konstantinos Kyprianidis

Automation in Energy and Environmental Engineering, Mälardalen University, Sweden
`esin.iplik@mdh.se`

Abstract

Several sensors are installed in the majority of chemical reactors and storage tanks to monitor temperature profiles for safety and decision-making processes such as heat demand or flow rate calculations. These sensors fail occasionally and generate erroneous measurement data that need to be detected and excluded from the calculations. However, due to the high number of process variables displayed in the chemical plants, this task is not trivial. In this work, a Bayesian network approach to detect faulty temperature sensors is proposed. By comparing the sensor measurements with each other, the faulty sensor is detected. A modular approach is preferred, and networks are created for 10 K temperature intervals to increase flexibility and sensitivity. Created networks can be adjusted for the operating temperature ranges; hence, they can be used for any catalyst and entire life cycle. The developed method is demonstrated on an industrial scale hydrocracker unit with 92 sensor couples installed in a series of reactors. From the investigated sensors, 16 of them showed a greater difference than the 2 K threshold chosen for the fault. In addition to that, 13 sensors showed an increasing temperature difference that may lead to a fault. Two scenarios were created to calculate the energy loss due to a faulty measurement, and a 5.5 K offset error was found to cause a 5.79 TJ energy loss every year for a small scale hydrocracker.

Keywords: Bayesian network, sensor fault, fault detection, diagnostics

Nomenclature

$P(a)$ Probability of an event a

$P(a \mid b)$ Probability of an event a given event b

BN   Bayesian Network

CPT   Conditional Probability Table

DAG   Directed Acyclic Graph

1 Introduction

Temperature is one of the most often measured process parameters for several industries. Reaction kinetics are highly dependent on the temperature of the process, making temperature measurements important for the chemi-

cal industry. These measurements are necessary to have knowledge of the reaction progress for both safety and control. Safety systems rely on the sensors, and wrong measurements can cause a false alarm or dangerous situations. According to the French Ministry of Ecology, Sustainable Development and Energy, sensor failures are responsible for 52% of the accidents in four most automated sectors (refining, metallurgy, food processing, and chemical processing) between 1993 and 2011 (Ministry of ecology, sustainable development and energy, 2012). Similarly, an optimal operation decision can only be guaranteed by measuring the relevant variables correctly. Faulty measurements do not necessarily cause accidents but waste energy and resources; therefore, sensor failure detection mechanisms are highly important for industrial plants.

Model-based and data-based methodologies are employed to detect sensor faults. While the first one compares the expected values calculated by using physical models with the measurements, the latter uses previous measurements to diagnose a failure. In literature, examples for both methods can be found for different applications. A mathematical model consisting of mass and energy balances was used for a central chilling plant's flow and temperature sensor fault detection (Wang and Wang, 2002). A nonlinear state observer with a mathematical model was tested to detect temperature sensor faults of a heat exchanger (Escobar et al., 2011). The model itself sets the limitation to the model-based approach because its accuracy defines the detection capability. It is suggested for the detection of total failures but not for small errors (Du et al., 2009). A data-based approach, an artificial neural network, was used to detect air flowrate sensor faults of a building (Wang and Chen, 2002). Later, a principal component analysis based fault detection system was developed for temperature sensors of a centrifuge chiller (Wang and Cui, 2005).

Bayesian networks (BN) have gained attention for fault detection as they can show causality. They are found similar to white-box models due to their qualitative graph aspect (Hommersom and Lucas, 2010). Although BNs need more information in terms of prior probabilities, they are used for industrial applications. For maintenance planning (Jones et al., 2010), root cause diagnosis of variations (Dey and Stori, 2005), prognosis (Hu et al., 2011) and risk analysis (Duval et al., 2012), these networks were found useful. Their use for fault detection was success-

fully shown on a chiller (Wang et al., 2017), a heat pump (Cai et al., 2014) and a fuel cell (Riascos et al., 2007). A sensor fault detection work was published for a medical body sensor network (Zhang et al., 2016).

The present paper shows a modular BN approach for fault detection of an industrial hydrocracker's temperature sensors. A modular approach is chosen for this case to cope with the increasing temperature of the system. The operational temperature range is divided into intervals, and a single network is trained for each interval. Further information on the developed strategy is given in the methods section. The results are presented and discussed in the section that follows, and a faulty sensor scenario is created to demonstrate the additional energy consumption due to the fault. Finally, in conclusion section outcomes of this work are summarized next to possible points for future investigations.

2 Methods

Bayesian networks are graphically represented probabilistic models, in which every node is assigned to a variable and has its conditional probability table (CPT) depending on its parents (Pearl, 1988). The graphical part of a BN is a directed acyclic graph (DAG). DAG is the qualitative element of a BN, which shows the causality relations between the variables. CPT is the quantitative element that carries the joint probability distribution knowledge, and the posterior probability is calculated according to Bayes's theorem (Eq. 1).

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \qquad (1)$$

In this work, Bayesian networks are trained for a fault detection system of temperature sensors in a hydrocracker unit. The sensors are distributed in concentric circles as given in Figure 1 in order to observe the temperature profile of the catalyst bed. There are four catalyst beds in the reactors, which have identical sensor positioning and all the sensors are thermocouples.

There is a pair of sensors at every point to acquire a reliable measurement, and the average value is used for process calculations. The continuous hydrocracking process suffers from catalyst degradation; due to this phenomenon, the temperature of the system has to increase over time to achieve the same level of cracking. The operation continues ideally without interruption until the end of the catalyst life. Measurements of five representative sensors over two years are given in Figure 2. Each sensor has an average 35 K increase in the measured temperature, and different catalyst beds have different operating temperatures. The measurements from all the sensors are in a 50 K range. Reference temperature (T_{ref}) is taken as the start of the run temperature.

A neighboring sensor pair with similar temperature measurements is selected for each location, and their average is used as a reference point to detect a faulty sensor of
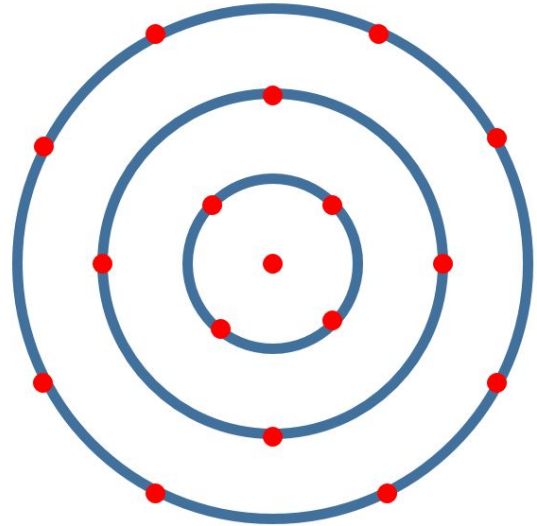
Figure 1. Approximate sensor locations on the horizontal cross-section.
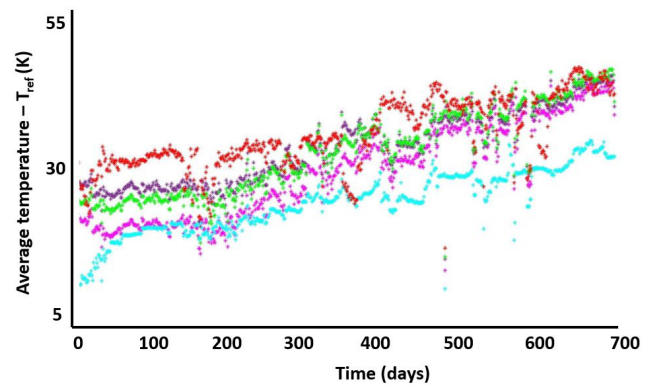
Figure 2. Daily data from five sensors for two years.

a pair. The average temperature of each sensor is plotted against the average neighbor temperature in Figure 3 to inspect the distribution of the data points in the operation domain.

As Figure 3 clearly demonstrates, there is a trend between the average values that shows well-chosen neighbors for each sensor; however, this trend is not entirely linear because of the flow regime, possible hot spots, and channels. If a single network were used to represent this space, the CPT would cover the entire area, most of which have zero probability. The operating temperature range of 50 K that is observed for an entire catalyst cycle is divided into 10 K fragments that are 5 K - 15 K, 15 K - 25 K, 25 K - 35 K, 35 K - 45 K, 45 K - 55 K based on the average temperature of the pair to reduce the size of the CPT. These values are calculated with the subtraction of the T_{ref} from the real values. A discrete network is trained for each range separately by using the data within the range collected from every sensor. Figure 4 shows the range of five networks in different colors and the shaded
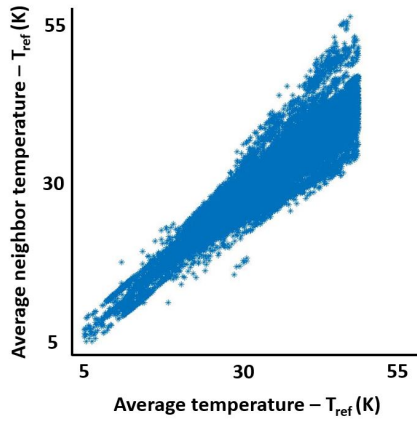
Figure 3. Average temperature range with respect to average neighbor temperature.

zero probability area, which proves the reduction of the CPT size.
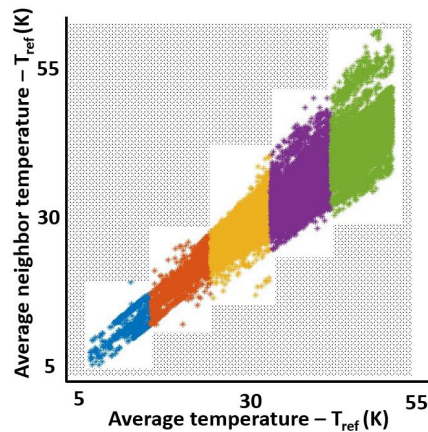
Figure 4. CPT reduction by modular network approach.

Five networks have identical DAG, which is given in Figure 5. The nodes a and b represent each sensor of a pair, and the neighbor node is the average value of the sensors selected for each pair differently. Structural learning of these three nodes is performed by using the necessary path condition algorithm (Steck and Tresp, 1999). Considering the sensitivity of the sensor type, a discretization interval of 1 K is chosen for a and b nodes, and 2 K for the neighbor node. Unlike the interval nodes of the sensors, the fault node and its CPT considering three states (No fault, Fault A, and Fault B) are added with expert knowledge, an exemplar CPT is given in the Table 1. For every neighbor, A and B interval, probability of fault states are defined. The fault node's CPT is a 3D array due to three parent nodes it has, and a high temperature difference of A and B indicates a fault. For example, the temperatures of Neighbor = T, A = T + 0.5, and B = T + 3.5 would result in a 95% probability of 'Fault B' because the difference is higher than the threshold, and the measurement of sensor A is closer to the neighbor measurement. Hugin, a
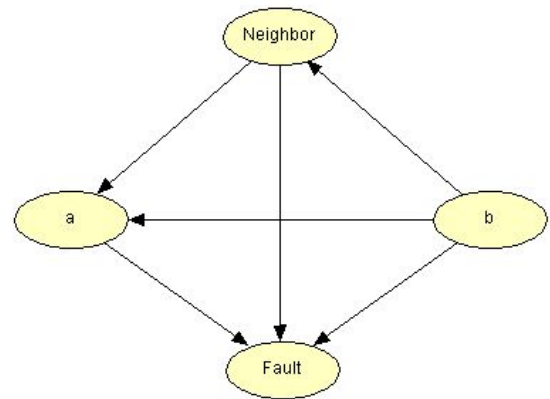
Figure 5. DAG of the Bayesian network used for sensor fault detection.

commercial software, is used to build BNs of this work.

Different hydrocracking catalysts operate in different temperature ranges. The suggested modular approach brings flexibility to sensor fault detection. A different range can easily be covered by building a new module with the same DAG and a similar CPT; therefore, the networks are easily adjustable and reusable for different catalysts and units. Although a single network for the entire operation domain might be useful for higher errors due to its CPT range, it cannot be reused in a different temperature range. Furthermore, a greater CPT range also causes problems. It is hard to fill with expert knowledge if it has the same discretization intervals, or it is not as sensitive if it has the same size as the smaller range CPT. If a single network were designed for the same system with equal intervals, the fault node's CPT size would be 187500. In the current design, each fault node has a CPT with 2700 values, and due to the similar trend, they are identical.

A single measurement per day is taken from 92 sensors for two years (684 from each sensor), and they are classified into five intervals of 10 K as described above. The outliers that are values outside the operational range due to some process interruptions are visually excluded from data sets. The five networks are built by using 767, 2494, 10909, 17437, and 17889 points according to increasing average temperature intervals. These networks are tested by using five test data sets of 709, 2436, 10791, 17642, and 17658 values, respectively, 684 measurements from each sensor, excluding the outliers. The test data points are chosen 12 hours after each training data; therefore, the testing is as independent as possible. Both the training and the test data covers the entire catalyst life span.

3 Results and Discussion

The five networks for different temperature intervals are built and tested with real data. According to interval cal-

Table 1. Construction rules of fault node CPT.

Neighbor	(T-1) – (T+1)						
A	(T) – (T+1)						⋯
B	(T) – (T+1)	(T+1) – (T+2)	(T+2) – (T+3)	(T+3) – (T+4)	(T+4) – (T+5)	(T+5) – (T+6)	⋯
No Fault	1	1	0.7	0	0	0	⋯
Fault A	0	0	0.05	0.05	0.05	0.05	⋯
Fault B	0	0	0.25	0.95	0.95	0.95	⋯

culus, the built CPT detects an error if there is a difference of at least 2 K. A deviation over 2 K between the pair is detected at 16 of 92 sensors. Each detected fault is investigated separately and found that the difference between the sensor measurements does not change over time. In Figure 6, error with respect to time is given for a pair of sensors.
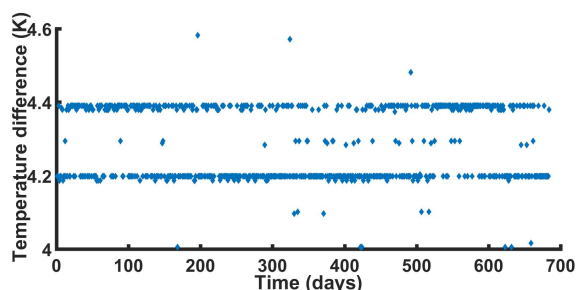
Figure 6. Temperature difference of a sensor pair, indicating a possible calibration error.

The constant temperature difference over time shows a possible calibration error starting from the installation. The smallest difference detected as calibration error is 2.1 K, and the greatest one is 5.5 K. This type of error may seem harmless since it does not increase over time; however, it can interfere easily with the calculation and application of optimal operating conditions. Optimality cannot be guaranteed with measurement errors existing in a system. If a calibration error is detected, it can be eliminated easily.

A calibration error is not the only error type found; an increasing deviation is also observed in faulty sensors. In Figure 7, an example of increasing deviation is given.
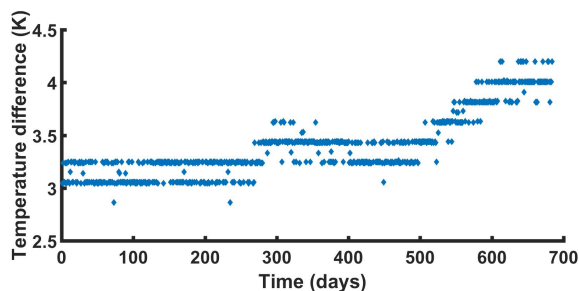
Figure 7. Increasing temperature difference of a sensor pair, indicating an error.

Besides the faulty sensors, 13 sensors are detected with

decreasing reliability, falling into 30% fault probability regions of the CPTs. These sensors have no calibration error and have an increasing temperature difference trend. Figure 8 shows the temperature difference increasing over time, which does not pass the error threshold of 2 K. If the same equipment is to be used for the next run of two years, these sensors should be carefully monitored.
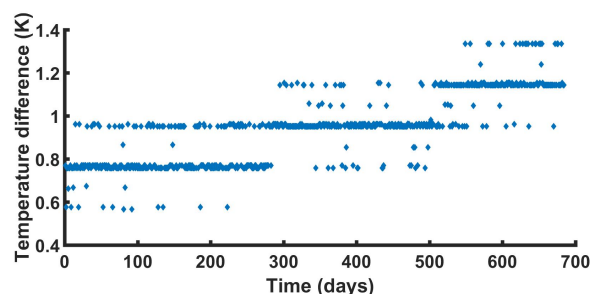
Figure 8. Increasing temperature difference of a healthy sensor pair with decreasing reliability.

Considering the lifetime of sensors are generally longer than two years, this information can be useful when there is an opportunity for maintenance. When there is a need for a catalyst change, the continuous operation stops for a certain amount of time, and all the reactor content is discharged. The internal equipment maintenance is often scheduled for this period. With the knowledge of calibration error and increasing deviation, relevant sensors can be fixed. The collected data belongs to a new set of sensors; therefore, it is expected to have a low rate of faults since sensor faults have a bathtub curve distribution over time (Mishra et al., 2002). A bathtub curve is given in Figure 9, which suggests increasing the importance of a fault detection mechanism with the aging equipment. A fault detection system can simplify maintenance planning because it allows an analysis of the given curve. Considering the age of the sensors, the expected lifetime of the next catalyst, and the number of faulty or less reliable sensors, a decision on whether to keep or change the equipment can be taken. A decision to reuse the same sensor system for the next two years' run can save the equipment cost; however, an erroneous measurement can increase operating costs.

Tolerance value of a thermocouple is around ±1 K and this value is important to test the robustness of the networks. Random noise is generated within this range and added to each test measurement. From 49236 measure-
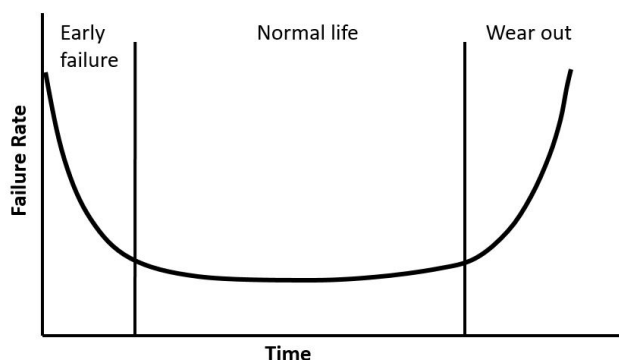
Figure 9. Bathtub curve of sensor failure rate with respect to time.

ments, 3.1% are detected false positive and 3.8% are detected false negative with the added noise.

Energy loss due to faulty sensors

Two scenarios are developed to calculate the energy loss due to measurement errors, the first one for a faulty sensor, and the second one for a 30% fault probability sensor with an increasing temperature difference trend.

Scenario 1

A small scale hydrocracker unit with 85000 kg/h feed capacity is fed with heavy oil of approximately 30 degrees API (at 60^oF) and 2.95 KJ/kg.K specific heat (United States Department of Commerce, Bureau of Standards, 1929). The start of the run temperature of the unit is 650 K, which is an average value for this operation (Ancheyta et al., 2005). A calibration error of 5.5 K decreases the average temperature of the pair by 2.75 K. If this sensor is a feed temperature measurement sensor, which is often a manipulated variable of the control system (Cutler and Hawkins, 1987), this error costs an additional 5.79 TJ energy consumption in a year for a constant temperature operation. As explained earlier, in reality, the system temperature increases as the severity of catalyst degradation increases; therefore, the specific heat increases over time, and the energy loss is higher than the calculated value.

Scenario 2

The same hydrocracker unit sensor has an increasing temperature deviation, which does not exceed the fault threshold yet, but it is in the 30% fault region. Its temperature difference increases by 0.2 K every four months, starting from 1 K to 2 K at the end of two years. It decreases the average temperature by 0.5 K in the first four months, causes an additional 0.361 TJ consumption. In the next five periods of four months, this value increases to 0.433 TJ, 0.505 TJ, 0.578 TJ, 0.65 TJ, and 0.722 TJ. In total, it might cause an additional consumption of 3.25 TJ in two years.

These two scenarios show why sensor faults need to be detected as early as possible. They both consider the fault in a negative direction. A fault that increases the average

temperature can cause other problems, for example, inefficient cracking. In the case of the lower temperature of the reactor feed, the cracking reactions cannot be carried out at the desired rate, and the product specifications cannot be reached. Therefore, the outputs of the process need further processing, and so additional energy need arises that increases the production cost.

An important point to consider is the limitations of the designed network system. As mentioned earlier, this type of limited range network cannot identify the significant errors, which throws the average out of the operational domain. It can be overcome by using high range buffer intervals as the first and the last elements of the CPT. This addition allows the BN to function in a higher range without expanding it significantly. As the targeted errors are high in value, the buffer zones do not need high sensitivity, so it is easy to detect them if the network can be evaluated with the given measurements. Another issue to be considered is the reliability of the neighbor sensors' measurements. If there is a faulty neighbor, it will still be possible to detect a fault, but it will not be possible to detect which sensor is faulty. A weighted average of multiple measurements can be used as the reference point to reduce this problem.

4   Conclusion and Future Work

Five Bayesian networks, each addressing a different temperature interval, are designed to detect a hydrocracker reactor's sensor faults. A set of data collected from 92 sensor pairs and their closest neighbors is used for training the networks, and a fault node is added with expert knowledge. Developed networks are tested with data, and 16 sensors are detected with an error greater than 2 K. Detected faulty sensors are investigated, and two types of faults are found in the system, offset errors and increasing errors. In addition to the faulty sensors, 13 sensors are detected with decreasing reliability, although their differences are lower than the fault threshold of 2 K. Two energy loss scenarios are used to demonstrate the importance of detecting both the erroneous measurements and the small deviations. BN is a fast tool that can be built both from data and by expert knowledge that gives it flexibility, and in this work, it is demonstrated that BNs can be used to detect sensor failure. With the increasing interest in it, BNs might help different industries operate safely without energy losses. Future work on the same system with special attention on the sensors that have an increasing temperature deviation can justify the reliability of this prognosis. The limitations discussed, significant errors' detection and neighbor faults, should be solved to build a reliable sensor fault diagnostics system. In addition to that, a dynamic Bayesian network, which considers the previous error, can be used to distinguish the increasing errors from the offset errors.

References

Jorge Ancheyta, Sergio Sánchez, and Miguel A Rodríguez. Kinetic modeling of hydrocracking of heavy oil fractions: A review. *Catalysis Today*, 109(1-4):76–92, 2005.

Baoping Cai, Yonghong Liu, Qian Fan, Yunwei Zhang, Zengkai Liu, Shilin Yu, and Renjie Ji. Multi-source information fusion based fault diagnosis of ground-source heat pump using Bayesian network. *Applied energy*, 114:1–9, 2014.

CR Cutler and RB Hawkins. Constrained multivariable control of a hydrocracker reactor. In *1987 American Control Conference*, pages 1014–1020. IEEE, 1987.

Story Dey and JA Stori. A Bayesian network approach to root cause diagnosis of process variations. *International Journal of Machine Tools and Manufacture*, 45(1):75–91, 2005.

Zhimin Du, Xinqiao Jin, and Yunyu Yang. Fault diagnosis for temperature, flow rate and pressure sensors in vav systems using wavelet neural network. *Applied energy*, 86(9):1624–1631, 2009.

Carole Duval, Geoffrey Fallet-Fidry, Benoît Iung, Philippe Weber, and Eric Levrat. A Bayesian network-based integrated risk analysis approach for industrial systems: application to heat sink system and prospects development. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 226(5):488–507, 2012.

RF Escobar, CM Astorga-Zaragoza, AC Téllez-Anguiano, D Juárez-Romero, JA Hernández, and GV Guerrero-Ramírez. Sensor fault detection and isolation via high-gain observers: Application to a double-pipe heat exchanger. *ISA transactions*, 50(3):480–486, 2011.

Arjen Hommersom and Peter JF Lucas. Using Bayesian net-works in an industrial setting: Making printing systems adap-tive. In *ECAI*, pages 401–406, 2010.

Jinqiu Hu, Laibin Zhang, Lin Ma, and Wei Liang. An integrated safety prognosis model for complex system based on dynamic Bayesian network and ant colony algorithm. *Expert Systems with Applications*, 38(3):1431–1446, 2011.

B Jones, Ian Jenkinson, Zaili Yang, and Jin Wang. The use of Bayesian network modelling for maintenance planning in a manufacturing industry. *Reliability Engineering & System Safety*, 95(3):267–277, 2010.

Ministry of ecology, sustainable development and energy. Accident analysis of industrial automation. Technical report, 2012.

Satchidananda Mishra, Michael Pecht, and Douglas L Goodman. In-situ sensors for product reliability monitoring. In *Design, Test, Integration, and Packaging of MEMS/MOEMS*, volume 4755, pages 10–19. International Society for Optics and Photonics, 2002.

Judea Pearl. Probabilistic reasoning in intelligent systems. *San Mateo, CA: Kaufmann*, 23:33–34, 1988.

Luis Alberto M Riascos, Marcelo G Simoes, and Paulo E Miyagi. A Bayesian network fault diagnostic system for pro-ton exchange membrane fuel cells. *Journal of power sources*, 165(1):267–278, 2007.

Harald Steck and Volker Tresp. Bayesian belief networks for data mining. In *Proceedings of the 2. Workshop on Data Mining und Data Warehousing als Grundlage moderner entscheidungsunterstützender Systeme*, pages 145–154. Citeseer, 1999.

United States Department of Commerce, Bureau of Standards. Thermal properties of petroleum products. Technical report, 1929.

Shengwei Wang and Youming Chen. Fault-tolerant control for outdoor ventilation air flow rate in buildings based on neural network. *Building and Environment*, 37(7):691–704, 2002.

Shengwei Wang and Jingtan Cui. Sensor-fault detection, diagnosis and estimation for centrifugal chiller systems using principal-component analysis method. *Applied Energy*, 82(3):197–213, 2005.

Shengwei Wang and Jin-Bo Wang. Robust sensor fault diagnosis and validation in hvac systems. *Transactions of the Institute of Measurement and Control*, 24(3):231–262, 2002.

Zhanwei Wang, Zhiwei Wang, Suowei He, Xiaowei Gu, and Zeng Feng Yan. Fault detection and diagnosis of chillers using Bayesian network merged distance rejection and multi-source non-sensor information. *Applied energy*, 188:200–214, 2017.

Haibin Zhang, Jiajia Liu, and Nei Kato. Threshold tuning-based wearable snsor fault detection for reliable medical monitoring using Bayesian network model. *IEEE Systems Journal*, 12(2):1886–1896, 2016.

Testing ERP and MES with Digital Twins

Juha Hirvonen   Hannu Hakalahti   Mikko Ylihärsilä

School of Technology, Seinäjoki University of Applied Sciences, Finland,
{juha.hirvonen, hannu.hakalahti, mikko.yliharsila}@seamk.fi

Abstract

Enterprise resource planning (ERP) systems and manufacturing execution systems (MES) are becoming more and more important also for the small and medium-sized enterprises (SMEs). Even though the failure rates of the ERP projects seem to be exaggerated, the failures and problems in the integration cost lots of time and money. Therefore, there is clearly a need for test environments for the ERP and MES systems. This paper presents an approach to test these information systems by connecting them to the simulation models of the production and thus generating a digital twin of the production and the ERP/MES systems. As a proof-of-concept, two different twins are constructed with two different software.

Keywords: digital twin, simulation, ERP, MES

1 Introduction

Enterprise resource planning (ERP) systems are integrated management systems of main business processes, such as production, input and output logistics, finance and accounting, sales and marketing, and human resources. An ERP system helps different parts of the organization to efficiently share data and information thus reducing costs and improving the management of business processes. Manufacturing execution systems (MES) are information systems for controlling, tracking, and documenting the process of manufacturing products it the factory floor. MES ensures effective production and provides information that can be utilized for improving the production output.

Traditionally, ERP and MES systems have been software for big companies. However, because of the ongoing era of digitalization and tightening competition, also small and medium-sized enterprises (SME) are more commonly investing in these systems (Supramaniam *et al*, 2014).

ERP projects are notorious for having low success rates and swollen costs, and there are numerous web pages by different consultant firms repeating the doomsday figures that assure that the frightening reputation is true. However, most of them do not have any reference for their figures and the ones that have seem to be cross-referencing each other or not-existing reports. Nevertheless, there are some actual publications by the consultant companies as well. According to a report by McKinsey, roughly 75 % of ERP projects fail to stay either on schedule or within budget, and around 67 % have a negative return on investment (Casanova *et al*, 2019). However, the report does not mention the exact source of these figures. Another report by Panorama Consulting Group (2020) states that 38 % of the ERP projects exceed the budget (average 66 % over) and 47 % do not stay in the schedule (average 33 % overtime). The report is based on a survey with 181 respondents, out of which 34 % are manufacturers (Panorama Consulting Group, 2020). Yet another report made by Mint Jutras (2019) states that 46 % of the ERP projects exceed the budget and 38 % do not stay in the schedule. It also states that 67 % rate their implementation as successful. The report is based on a survey of 315 respondents. These figures are much more positive than the ones reported by e.g. McKinsey and fairly in line with the ones presented by Panorama Consulting Group. However, the authors also suspect that based on the actual benefits realized, many of the respondents might be overrating their success (Mint Jutras, 2019).

Some recent academic research has been conducted on the success of the ERP projects as well. Johansson and Sudzina (2019) studied the ERP implementations in the European SMEs and they found that only 26 % exceeded the budget, of which around one third exceeded the budget with 50 % or more. The survey had 121 respondents from Denmark, Slovakia and Slovenia (Johansson and Sudzina, 2019). As stated by Haddara and Zach, (2012), the ERP implementation projects are typically easier and shorter in SMEs since the size of the organization is smaller and the business complexity lower. However, the staff must often play multiple roles in small companies, and thus allocating enough resources to the ERP project might be problematic (Mint Jutra, 2019). Also, for the SME, the overruns in the budget and time can be even more disastrous than for a big company.

Based on the surveys presented, the failure rates of the ERP projects seem to be overrated. Nevertheless, the budget and schedule overruns are quite common. Also, the unchallenged fact is that some extremely expensive ERP implementation catastrophes have occurred (e.g. listed in Ram *et al*, 2013), and many times the employees have found the fresh ERP system to be too

stiff and to lead to some inefficient practices such as process duplication (Kerr and Houghton, 2014). To support the implementation process and to estimate the possible bottlenecks, different tools and analyses such as critical success factors (Ram *et al*, 2013), critical success strategies (Yeh and Xu, 2013), critical failure factors (Amid *et al*, 2012), project risk modeling (Aloini *et al*, 2012), and maturity models (Kerr and Houghton, 2014) have been developed. However, these approaches may be too theoretical and too complex especially for the needs of the SMEs. More concrete testing procedures would be beneficial.

The actual testing of the ERP and MES systems in the production environment before the investment decision is extremely difficult. Often, the only option is to trust in the supplier's promises. In this paper, a testing approach based on digital twins is suggested.

Barricelli *et al* (2019) define digital twins as computer-based models that are simulating, emulating, or mirroring a physical entity or a process. They emphasize that instead of "just" a simulation, it is the virtual counterpart of the entity or the process, and the data flow in and out of the twin is essential. Manufacturing is one of the main application domains of the digital twins, and the application cases include factory design, manufacturing optimization, consistent process monitoring, and predictive maintenance (Barricelli *et al*, 2019). Thus, it seems a promising technology also for testing the IT systems for the manufacturing companies.

In this paper, a proof-of-concept of using digital twins to test ERP and MES systems is presented. The main target group of the approach is SMEs. The digital twins are constructed by generating the simulation models of two different manufacturing systems and connecting them to the custom MES systems and an ERP system with an online shop interface. This way, handling the customer order, starting the production, monitoring the production process, and receiving data to the IT systems from the production can be experimented.

The rest of the paper is organized as follows: Section 2 discusses the pilot production environments used in this study, Section 3 presents the digital twins produced, and Section 4 draws the conclusion.

2 Pilot Production Environments

This section presents the two manufacturing systems that were used as pilots in this study. They were the educational cyber-physical factory and a flexible manufacturing system (FMS) cell. The systems are presented in their sub-sections.

2.1 Educational Cyber-Physical Factory

The cyber-physical factory is a modular production line for education purposes made by Festo Didactic. The factory assembles simple cell-phone-like products

consisting of a back cover, a printed circuit board (PCB) and fuses. The PCBs are made of plastic and they do not have conductive tracks, pads or connectors – only the holders for the fuses. The purpose of the cyber-physical Factory is to demonstrate different assembly processes and hardware as well as offer insight into the modular design, networked logistics, and MES integration. Altogether 192 different product variants can be assembled from the components of different colors.

The cyber-physical factory consists of a high-bay storage for the back covers and the assembled products, a drilling station, a robot assembly cell, in which the robot mounts the PCBs to the back covers and the fuses to the PCBs, and a machine-vision-based inspection station for assuring the correct composition of the assembled product. The factory is controlled by Festo's own MES4 software. A custom MES software made by the students is in use as well. Figure 1 presents the cyber-physical factory, and Figure 2 shows an example of a product that can be assembled with the factory.
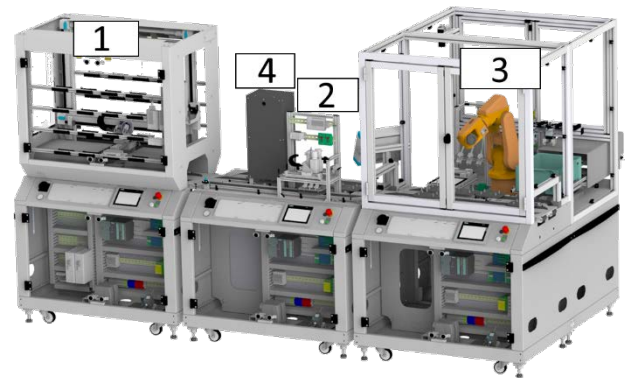
Figure 1. The educational cyber-physical factory. The high-bay rack (1), the drilling station (2), the robot assembly cell (3), and the machine vision station (4). The modules are numbered in the order of the process flow.
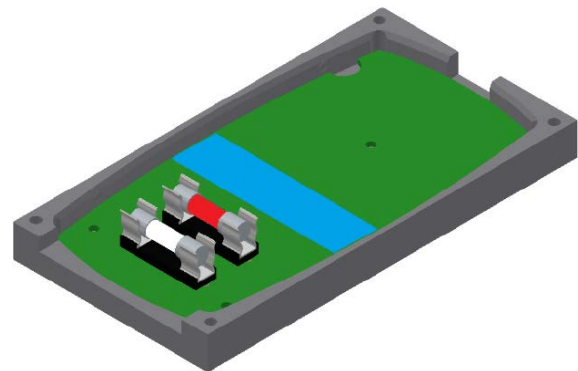
Figure 2. One product variant of the cyber-physical factory consisting of a gray back cover, a blue PCB and transparent and red fuses.
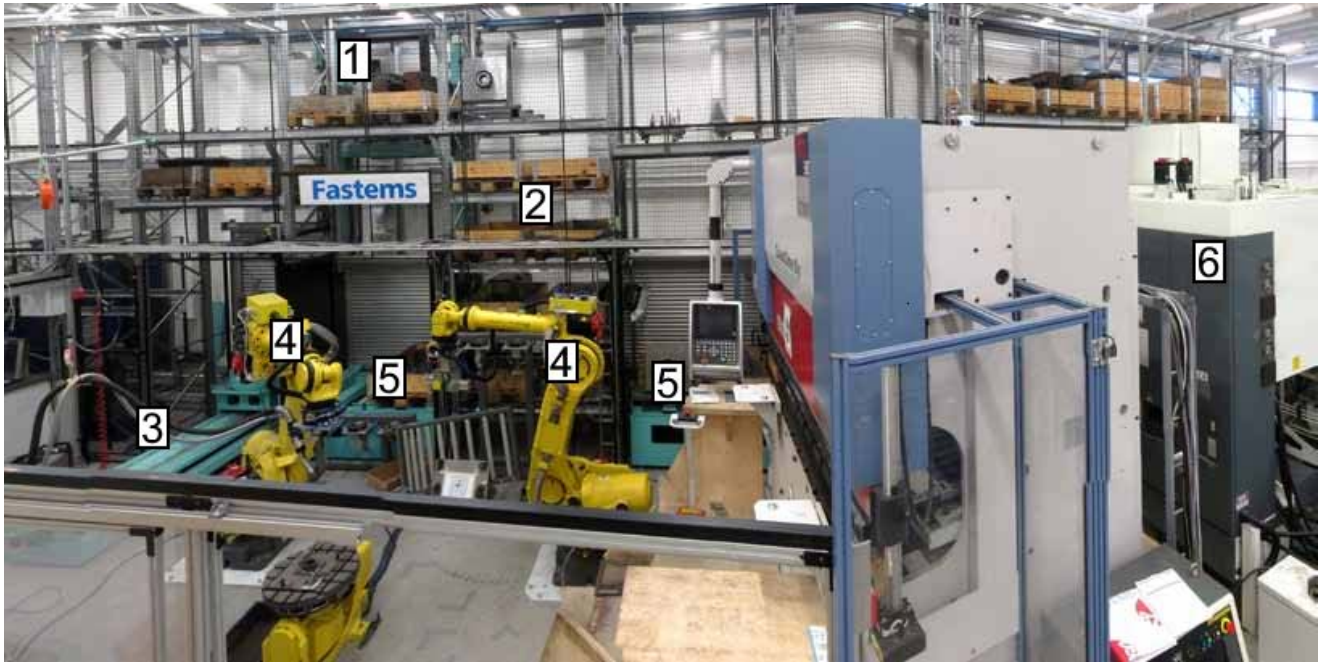
Figure 3. The FMS cell. The stacker crane (1), the pallet storage (2), the loading station (3), the industrial robot (4), the pallet stations (5), and the machining center (6).

2.2 FMS Cell

The flexible manufacturing system (FMS) cell by Fastems consists of a stacker crane, a pallet storage for wooden Euro-pallets and machining pallets, a loading station for the machining pallet, an industrial robot, pallet stations inside the robot cell, and a machining center. The MES system used to control the production of the cell is Fastems MMS. The FMS cell can be configured to produce a high variety of products, and the cell is manufacturing support parts in the example application of this study. The production cycle has six phases, and it goes as follows: (1) The stacker crane picks a pallet containing the blank and an empty pallet from the pallet storage and places them into the pallet stations of the robot. (2) The robot picks the blank from the pallet in the pallet station and places it to the machining pallet in the loading station. (3) The stacker crane takes the machining pallet and the blank to the machining center, where the support part is machined from the blank. (4) The stacker crane returns the machining pallet and the machined support part to the loading station after machining. (5) The robot picks the support part from the machining pallet and places it to the machine vision inspection. (6) The robot places the support part to the pallet for the machined parts after the inspection. Figure 3 shows the FMS cell.

3 Digital Twins

This section presents the digital twins of the production environments discussed in the previous section. The ERP system used with both the digital twins was Odoo. Odoo is an open-source ERP system and it was chosen because it is attractive for SMEs due to its affordable price (community version: free) and the high availability of different modules and add-ons that make the software scalable. The users can also create their own Odoo modules.

In order to send the manufacturing orders from Odoo to the simulation models, it was necessary to develop a small Odoo application for this purpose. In practice, the Odoo application reads the sale order data, creates a manufacturing order, and sends it to an external Python application.

3.1 Python web application

The Python web application acts as an intermediary between Odoo and the simulation models. It receives the manufacturing order from the Odoo, validates the order data, creates the steps of the work rotation, and sends commands to the simulation models. It also receives data from the simulation models. The application communicates with the simulation models via TCP/IP sockets and the data is sent and received in JSON format.

The application has a user interface in the form of a web page, where the user can inspect and get information from the states of the devices in the models, monitor the storage inventories, run the individual subprogram of the devices, create new manufacturing orders, and control (start/delete) pending manufacturing orders. Thus, the application can be considered as an MES tailored for the simulation models.

The application is implemented with Flask micro web framework and the web page uses Bootstrap 4 CSS framework for the visual appearance.
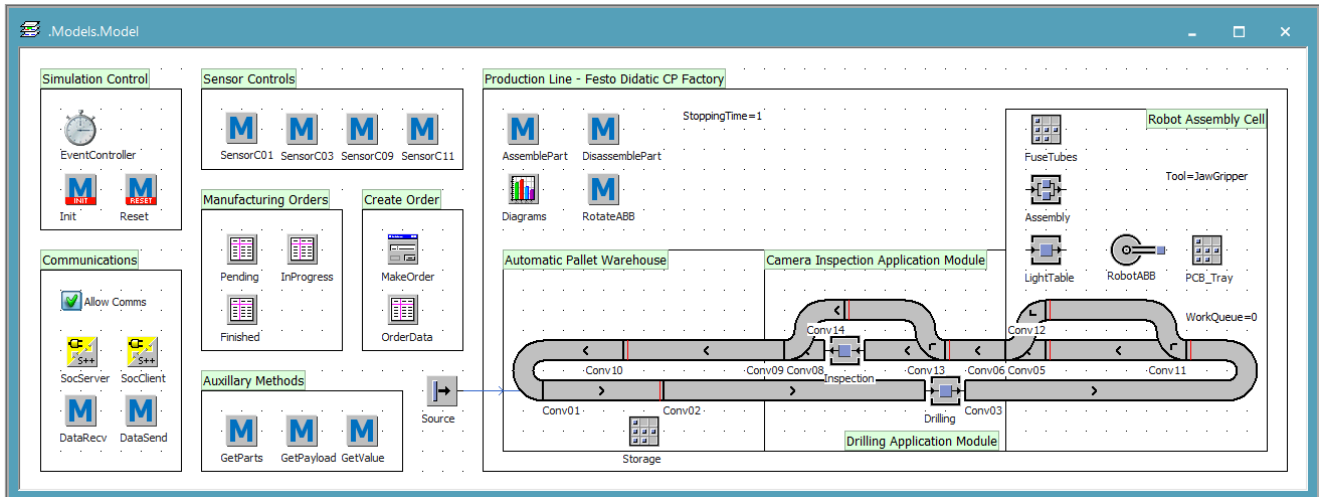
Figure 4. The digital twin of the production process of the cyber-physical factory.

3.2 Educational Cyber-Physical Factory

The simulation model of the Cyber-Physical Factory was made with Tecnomatix Plant Simulation, and the manufacturing orders are created with the Odoo ERP system. The product variants are generated in the Siemens Teamcenter PDM system from where a Python accessory inserts them into the Odoo database. In the database, a bill of materials (BOM) is created for the product and the product is published in the Odoo online shop. In the online shop, the customer can choose the colors (and in the case of the fuses, the number) of the components and make the order.

The Plant Simulation model assembles the product as well as sends information about the state of the assembly process and the current storage inventories back to the web application. Figure 4 presents the digital twin of the production process of the cyber-physical factory, and Figure 5 shows the view from the Odoo online shop.

3.3 FMS Cell

The simulation model of the FMS cell was made with Visual Components. The industrial robots, workpiece positioner, Euro-pallets and shelves were added from the model library of the Visual Components while the rest of the devices and accessories were designed in Siemens NX. The CAD files were then imported into the Visual Components where the properties of the devices were defined.

In Visual Components, the functionality of the devices can be extended with scripts that are written in the Python 2.7 programming language. In the simulation model of the FMS cell, a Python script enabling a TCP/IP connection to an external Python web application was added to the stacker crane, the loading station, the robot and the machining center. Thus, these devices can send and receive data.
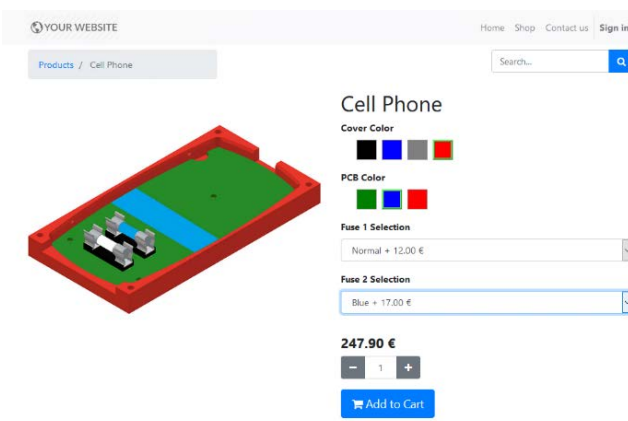
Figure 5. View from the Odoo online shop for ordering the "cell phones".
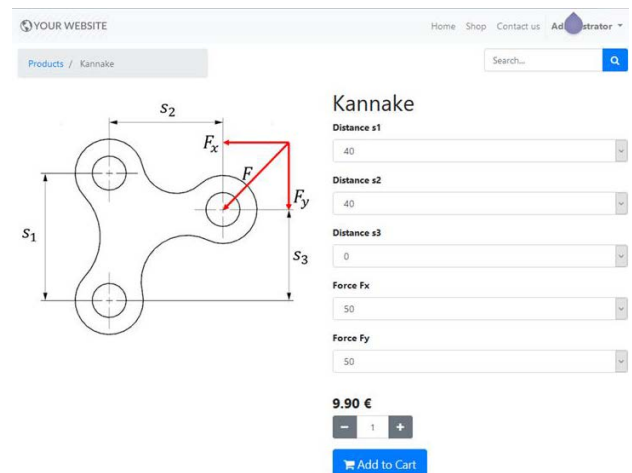
Figure 6. View from the Odoo online shop for ordering the support parts.

Figure 4. The digital twin of the FMS cell.

The simulation model of the FMS cell can run two work rotations: machining of a support part and polishing a cube. In the Odoo online shop, the customer can set the distances of the holes of the support, and the X and Y components of the force acting to the support. The machining time of the support is calculated based on these parameters. During the machining work rotation, the machining time is sent to the machining center.

Figure 6 shows the view from the online shop, and the digital twin of the FMS cell is seen in Figure 7.

4   Results and Discussion

The digital twins of the educational cyber-physical factory and the FMS cell were built and connected with the Odoo ERP system as discussed in Chapter 3. Also, the physical cyber-physical lab was connected with the Odoo ERP. Placing customer orders in the online shop as well as accepting the orders and starting the production in the admin side are functioning similarly with the physical system and the digital twin. The FMS cell will also be connected with Odoo in the near future. So far, the tests have been limited to creating individual device interfaces to the industrial robot and the machining center. The simulation models mimic the work cycles of their physical counterparts, but all the details were not modeled – for example, the shape of the ready support part in the digital twin of the FMS cell is always similar and do not imitate the ordered part. However, all the tests described here offer a proof-of-concept of the approach.

Apart from testing the ERP–MES integration, this approach could be utilized to test the capacity to produce different kinds of big orders and estimate the delivery time. Also, these kinds of digital twins can be used widely for educational purposes. This is resource efficient as a big group of students can work simultaneously with the models instead of working in shifts in one laboratory. The model of the FMS cell presented in this paper has already been utilized in teaching e.g. socket programming. Currently, the simulation models lack the functionality to inform the MES or ERP of an error that occurs in the production, but this kind of functionality can be added. This addition would give also valuable information about the data transfer from the error situations of the production to the information systems.

One important aspect is the modularity of the presented approach. This is estimated below separately for both twins.

The digital twin of the educational cyber-physical factory consisted of the Odoo online shop, the PDM-Odoo interface, the Odoo-Plant Simulation interface and the Plant Simulation model. If some company would like to test the suitability of Odoo to their production, the Plant Simulation model and the PDM-Odoo interface should be fully reconstructed. However, the PDM-Odoo interface is not compulsory as the product variants can be created in Odoo itself in the new versions (Odoo 13 or above). The Odoo-Plant Simulation interface would need to be updated but much of it could be reused. On the other hand, if a different ERP system should be tested with the re-existing model, the ERP-Plant Simulation interface should be reconstructed.

The digital twin of the FMS cell consisted of Odoo online shop, Odoo-Visual Components interface (the

REST interface), Visual Components model and the TCP/IP socket communication blocks inside the model. Again, if the production scheme changes, the Visual Components model needs to be reconstructed. The Odoo-Visual Components interface would need an update but many of its parts would be still usable. The socket communication blocks would need no more than a minor update. Changing the ERP system would need reconstructing the ERP-Visual Components interface. It is also worth noting, that the Odoo-Plant Simulation interface and the Odoo-Visual Components interface are mostly similar instead of few conditional statements. The major reason for the differences in the interfaces is the difference in functioning of the physical devices: the modules of the educational cyber-physical factory poll actively the MES whereas the devices of the FMS cell just wait for the orders from the MES.

5 Conclusion

This paper presented the concept of testing ERP and MES integration with a digital twin approach. The benefit of the approach is that it enables testing the ERP-MES integration beforehand without the need to interrupt the actual production. The approach gives a concrete demonstration of the functioning of the ERP system. Therefore, making the order in a online shop, handling orders, and the exchange of the data between the systems can be tested with it. Two different simulation models were created with two different modeling software to show the flexibility of the method.

The main downside of the approach is that generating a digital twin takes time. If the production system of the company is very complex, then a highly complex model must be constructed as well. However, as the target group of the method is SMEs, the production is not expected to be extremely complicated. Another disadvantage is that the interfaces between the model and the ERP/MES systems should be tailored for each ERP system. Yet, the selection of the ERP systems suitable for SMEs is not that large and thus this problem is not that crucial. Lastly, the constant question with the models is how well they correspond to the actual system. Precision is needed in generating the digital twins.

Although there are some open questions, the demonstrations of the digital twin approach to test ERP and MES are really promising. Another ERP system and some commercial MES systems will be tested in the near future. Future work includes also adding more intelligence to ordering and manufacturing. Firstly, the support part orders made for the FMS system will be passed to a CAD program and further to FEM calculation to prevent the user from ordering a physically inconsistent part. Furthermore, the geometry of the support part will be fed into CAM software for automatic CNC cutting program generation. The generated CNC program will be simulated to validate for collision-free path and calculate the part's machining time. Calculated machining time will be further used to fine-tune the full factory simulation times, calculating more precise cost estimations and price quotations.

References

Davide Aloini, Riccardo Dulmin, and Valeria Mininno. Modelling and Assessing ERP Project Risks: A Petri Net Approach. *European Journal of Operational Research*, 220(2):484–495, 2012. doi: 10.1016/j.ejor.2012.01.062.

Amin Amid, Morteza Moalagh, and Ahad Zare Ravasan. Identification and Classification of ERP Critical Failure Factors in Iranian Industries. *Information Systems*, 37(3):227–237, 2012. doi: 10.1016/j.is.2011.10.010.

Barbara Rita Barricelli, Elena Casiraghi, and Daniela Fogli. A Survey on Digital Twin: Definitions, Characteristics, Applications, and Design Implications. *IEEE Access*, Vol 7, pp. 167653–167671, 2019. doi: 10.1109/ACCESS.2019.2953499.

Didier Casanova, Swati Lohiya, Jerome Loufrani, Matteo Pacca, and Peter Peters. Agile in Enterprise Resource Planning: A Myth No More. *McKinsey Digital*, 2019.

Moutaz Haddara, and Ondrej Zach. ERP Systems in SMEs: an Extended Literature Review. *International Journal of Information Science*, 2(6):106–116, 2012. doi: 10.5923/j.ijis.20120206.06.

Björn Johansson, and Frantisek Sudzina. Actual vs. Planned ERP Systems Implementation Costs in European SMEs. *Proceedings of the European Conference on Information Management and Evaluation*, pp. 232–238, 2019.

Don Kerr, and Luke Houghton. The Dark Side of ERP Implementations: Narratives of Domination, Confusion and Disruptive Ambiguity. *Prometheus*, 32(3):281–295, 2014. doi: 10.1080/08109028.2015.1017247.

Mint Jutras. The Real Facts about ERP Implementation: Busting the Myth of Failure, But Are You Overrating Your Success? Ultra Consultants, 2019.

Panorama Consulting Group. The 2020 ERP report, 2020.

Jiwat Ram, David Corkindale, and Ming-Lu Wu. Implementation Critical Success Factors (CSFs) for ERP: Do They Contribute to Implementation Success and Post-implementation Performance? *International Journal of Production Economics*, 144(1):157–74, 2013. doi: 10.1016/j.ijpe.2013.01.032.

Mahadevan Supramaniam, Azween Abdullah, and Ramachandran Ponnan. Cost Analysis on ERP System Implementation amongst Malaysian SMEs. *International Journal of Trade, Economics and Finance*, 5(1):72–76, 2014. doi: 10.7763/ijtef.2014.v5.343.

Chung-Hsing Yeh, and Yan Xu. Managing Critical Success Strategies for an Enterprise Resource Planning Project. *European Journal of Operational Research*, 230(3):604–614, 2013. doi: 10.1016/j.ejor.2013.04.032.

CFD Investigations of Subcooled Nucleate Boiling Flows and Acting Interfacial Forces in Concentric Pipes

Achref Rabhi[1]   Ioanna Aslanidou[1]   Konstantinos Kyprianidis[1]   Rebei Bel Fdhila[1,2]

[1]Future Energy Center, Mälardalen University, Sweden,
{achref.rabhi,ioanna.aslanido,konstantinos.kyprianidis,rebei.bel.fdhila}@mdh.se
[2]Hitachi ABB Power Grids, Sweden, rebei.bel_fdhila@hitachi-powergrids.com

Abstract

Boiling flows are widely encountered in several engineering and industrial processes. They have a special interest in nuclear industry, where a Computational Fluid Dynamic (CFD) thermohydraulic investigation becomes very popular for design and safety. Many attempts to model numerically subcooled nucleate boiling flows can be found in the literature, where several interfacial forces acting on bubbles which are interacting on the bulk fluid were neglected, due to the hard convergence of the calculations, or to the bad accuracy of the obtained results. In this paper, a sensitivity analysis is carried out for the interfacial forces acting on bubbles during subcooled nucleate boiling flows. For this purpose, 2D CFD axisymmetric simulations based on an Eulerian approach are performed. The developed models aim to mimic the subcooled nucleate boiling flows in concentric pipes, operating at high pressure. The predicted spatial fields of boiling quantities of interest are presented and commented. The numerical results are compared against the available experimental data, where it is shown that neglecting some interfacial forces like the lift or the wall lubrication forces will yield to good predictions for some quantities but will fail the prediction for others. The models leading to the best predictions are highlighted and proposed as recommendations for future CFD simulations of subcooled nucleate boiling flows.

Keywords: subcooled nucleate boiling flows, computational fluid dynamics, interfacial forces, sensitivity analysis

1 Introduction

Subcooled nucleate boiling flows are present in several engineering equipment and devices, such as nuclear reactors, heat exchangers and cooling systems. They are associated with hydrodynamics, heat and mass transfer. These flows have gained a special interest because of their enhanced heat transfer coefficients and improved heat transfer performance. Subcooled nucleate boiling flows are two-phase, composed by a continuous subcooled liquid phase, and a vapor dispersed phase, generated by separate bubbles. These bubbles are nucleated from micro-cavities, referred to as nucleation sites, randomly distributed over a heated surface, when the latter's temperature exceeds the working liquid saturation temperature at the local pressure. Bubbles depart their nucleation sites when reaching a critical size, referred as bubble departure diameter. They slide along the heated surface, and lift-off to the core of the liquid, where they migrate into the subcooled region of the flow. In this region, bubbles are subject to condensation when the local flow temperature is lower than the saturation temperature. Hence, the two-phase boiling flow is referred as subcooled boiling flow.

In order to improve thermal management system design and to maximize heat transfer coefficients, high-fidelity models need to be incorporated. Experiments can be used successfully to fulfill this task. However, they are still very expansive and time consuming to perform. Eulerian two-fluid models based on CFD simulations are a promising solution to develop such high-fidelity representations, where the flow structure, transport phenomenon and interactions between phases can be well predicted. The vapor-averaged void fraction, width of two-phase layer near the heated surface and temperature fields, which are crucial for design and optimization, can be very well and accurately presented. Turbulence effects can also be incorporated through these models. One of the most successful Eulerian two-fluid model for boiling flows was developed by Kurul and Podowski (1990), where they decomposed the total applied heat flux at the wall on three components; the first accounts for the evaporation, the second accounts for the quenching contribution and the last is responsible for heating the liquid near the heated surface when bubbles are absent.

Based on the same framework approach, and applying some slight modifications, many authors simulated the subcooled nucleate boiling flows in different geometries at multiple operating conditions, and compared their numerical results to their experimental data, or data from literature (Lai and Farouk, 1993; Anglart and Nylund, 1996; Roy et al., 2001; Končar et al., 2004). However, no sensitivity analysis quantifying the effect of the used interfacial forces models was carried out, despite that the dispersed-phase spatial distribution is highly affected by these forces, in addition to the turbulent behavior of the flow, that will in turn significantly influence the hydrodynamics and the thermal process, including heat and mass transfer (Končar et al., 2004). Hence, the obtained results

were acceptable for some fields, but failed for many others. Also, the bubble's size in the bulk fluid was ignored, despite its importance for such flows and its major effect on heat transfer, temperature fields, void fraction distribution and the two-phase layer thickness.

A first model to predict the bubble size evolution in the bulk fluid was proposed by Ishii et al. (2005) via the Interfacial Area Transport Equation (IATE), where the authors proposed empirical model coefficients to take into account the coalescence and the turbulent break-up phenomenons. The IATE model and the proposed coefficients by Ishii et al. (2005) were adopted in CFD simulations by Končar and Krepper (2008) and Michta (2011) among others, for high pressure boiling in conventional sized channels, and by Rabhi and Bel Fdhila (2019) for low pressure boiling in narrow channels. However, the bubbles bulk diameter was not very well predicted, and affected the accuracy of the computed fields. This is due to the used model coefficients which are not applicable for different geometries and operating conditions. In addition, the IATE model depends on the used interfacial forces, but, as highlighted previously, a sensitivity analysis and a quantification of these forces are missing, and represents a research gap for boiling flows modeling. However, assuming a constant bubble bulk diameter model, set to the mean value of the experimentally measured bubble bulk diameter, represents a good approximation, and yields to acceptable predictions as it will be shown in this work.

In the present work, 2D axisymmetric CFD simulations based on an Eulerian two-fluid approach are carried out modeling the subcooled nucleate boiling flow of refrigerant R-113 flowing upward a concentric pipe at an absolute pressure of 2.69 bar. The CFD model is implemented on the open-source CFD platform OpenFOAM, where a Finite-Volume discretization of the governing transport equations is adopted. The spatial distributions of boiling quantities predicted by the CFD calculations are presented and analyzed. The effect of the interfacial forces on the predicted fields are investigated. The model predictions are compared against the available experimental data of Roy et al. (2001), and these models yielding to successful predictions are highlighted and proposed as recommendations for future subcooled nucleate boiling flows CFD simulations.

2 Mathematical model

Based on the Eulerian two-fluid framework, transport equations, i.e., mass, momentum, energy and turbulence, are solved for the continuous liquid phase and the dispersed vapor phase.

2.1 Transport equations

Neglecting the small density variation with temperature rise, for each phase k, the mass conservation transport equation is written as:

$$\frac{\partial \alpha_k \rho_k}{\partial t} + \nabla \cdot (\alpha_k \rho_k \mathbf{U}_k) = \Gamma_{ki} - \Gamma_{ik} \qquad (1)$$

where α is the void fraction, ρ is the fluid density, $\mathbf{U}$ is the velocity vector and Γ is mass transfer rate, that denotes mass transfer rate per unit volume due to evaporation from the liquid phase to the vapor phase or mass transfer rate per unit volume by condensation from the vapor phase to the liquid phase, that their expression will be given in the boiling model section.

The momentum transport equation for each phase is given by:

$$\frac{\partial \alpha_k \rho_k \mathbf{U}_k}{\partial t} + \nabla \cdot (\alpha_k \rho_k \mathbf{U}_k \mathbf{U}_k) = -\alpha_k \nabla p + \mathbf{R}_k$$
$$+ \mathbf{M}_k + \alpha_k \rho_k \mathbf{g} + (\Gamma_{ki} \mathbf{U}_i - \Gamma_{ki} \mathbf{U}_k) \qquad (2)$$

where the first term in the right hand side corresponds to the pressure drop contribution, the second term $\mathbf{R}_k$ denotes the combined turbulent and viscous stress, the third term $\mathbf{M}_k$ represents the interfacial momentum transfer contribution, that will be detailed in the next section, the fourth term represents the gravity contribution and the last term accounts the contribution of mass transfer between phases to the momentum.

In this work, energy transport equation for each phase k is solved in terms of the specific enthalpy h. It is written for each phase as:

$$\frac{\partial \alpha_k \rho_k h_k}{\partial t} + \nabla \cdot (\alpha_k \rho_k \mathbf{U}_k h_k) = \alpha_k \frac{Dp}{Dt}$$
$$+ \nabla \left(\alpha_k D_{t,k}^{eff} \nabla h_k \right) + \Gamma_{ki} h_i - \Gamma_{ik} h_k + Q_{wall,k} \qquad (3)$$

where Dp/Dt is the material derivative of the pressure, $Q_{wall,k}$ is the product of the applied heat flux on the wall q_w'' and the contact area with the wall per unit volume, and $D_{t,k}^{eff}$ is the effective thermal diffusivity, that includes the turbulent contribution, given by:

$$D_{t,k}^{eff} = \frac{\lambda_k}{C_{p,k}} + \frac{\mu_k^t}{\Pr^t} \qquad (4)$$

where λ_k is the thermal conductivity, $C_{p,k}$ is the specific heat capacity, μ_k^t is the dynamic turbulent viscosity and $\Pr^t$ is the turbulent Prandtl number, set to a constant equal to 0.85.

To account the turbulent behavior of both phases, continuous and dispersed, the standard $k - \varepsilon$ turbulence model is used. For each phase, the turbulent kinetic energy k and its dissipation ε are calculated. The $k - \varepsilon$ turbulence model is adopted here because of its valid for fully turbulent flows, which is the case of the simulated problem here with a Reynolds number of 34450. It stills computationally cheap and robust turbulence model for pipes flows where there are no flow recirculation, despite its known limitations, rising principally from its empirical coefficients. However, this model lacks of sensitivity to adverse pressure gradients, and performs poorly for complex flows. In this work, the bubble induced turbulence was accounted by a turbulent kinematic viscosity as:

$$\nu_l^{t,b} = \frac{C_{\mu,b} d_b \alpha_v}{2} ||\mathbf{U}_v - \mathbf{U}_l|| \qquad (5)$$

where $C_{mu,b}$ is a coefficient set to 1.2 following the recommendations of Sato et al. (1981) for bubbly flow modeling, and d_b is the bubble bulk diameter, calculated via a constant bubble bulk diameter model. This bubble bulk diameter is set to the mean value of the experimental measured diameters by Roy et al. (2001).

2.2 Interfacial forces

For boiling flows, the relevant interfacial forces that should be included in the momentum equation are the drag, added mass, lift, wall lubrication and turbulent dispersion forces. In the following, the physical interpretation, mathematical and force coefficient models are presented.

Drag force

The drag force $\mathbf{F}_D$ models the resistance experienced by a bubble opposed to its motion, created by the surrounding fluid. Its mathematical expression is given by:

$$\mathbf{F}_D = -\frac{3}{4}C_D \frac{\alpha_v \rho_l}{d_b}\|\mathbf{U}_v - \mathbf{U}_l\|\left(\mathbf{U}_v - \mathbf{U}_l\right) \tag{6}$$

where C_D is the drag force coefficient. In this work, two different coefficient models are tested. The first model is the classic solid sphere drag model, proposed by Schiller and Naumann (1935) as:

$$C_D = \begin{cases} 24\left(1+0.15\mathrm{Re}_b^{0.687}\right) & \text{if } \mathrm{Re}_b \leq 1000 \\ 0 & \text{else.} \end{cases} \tag{7}$$

where Re_b is the bubble Reynolds number. The second is the model of Tomiyama et al. (1998), which was developed for bubbly flow. This model is written as:

$$C_D = \max\left(\frac{16}{\mathrm{Re}_b}\min\left(1+0.15\mathrm{Re}_b^{0.687},3\right),\frac{8}{3}\frac{\mathrm{Eo}}{\mathrm{Eo}+4}\right) \tag{8}$$

where Eo is the Eötvös number.

Added mass force

The acceleration of bubbles in the flow accelerates the surrounding fluid itself, which adds an apparent mass to the bubbles. This is modeled by the virtual mass force $\mathbf{F}_{VM}$, which is given by:

$$\mathbf{F}_{VM} = -C_{VM}\frac{1+2\alpha_v}{\alpha_l}\alpha_v\left(\frac{D\mathbf{U}_v}{Dt} - \frac{D\mathbf{U}_l}{Dt}\right) \tag{9}$$

where C_{VM} is the virtual mass coefficient set to a constant equal to 0.5 in this work following the recommendations of Zuber (1964) for spherical bubbles.

Lift force

The lift force $\mathbf{F}_L$ is caused by the velocity gradient across the dispersed phase. It is responsible for a pressure difference across the walls and creates a force towards the side. Its general mathematical expression is written as:

$$\mathbf{F}_L = C_L \alpha_v \alpha_l \left(\mathbf{U}_v - \mathbf{U}_l\right) \times \nabla \times \mathbf{U}_l \tag{10}$$

where C_L is the lift force coefficient. In this work, a constant lift force coefficient models are used, set to 0.05 and 0.5, in addition to the model of Tomiyama et al. (2002), proposed as following:

$$C_D = \begin{cases} \min\left(0.288\tanh\left(0.12\mathrm{Re}_b\right), f\left(\mathrm{Eo}\right)\right) & \text{if } \mathrm{Eo} \leq 4 \\ f\left(\mathrm{Eo}\right) & \text{if } 4 \leq \mathrm{Eo} \leq 10.7 \\ -0.288 & \text{else.} \end{cases} \tag{11}$$

with:

$$f\left(\mathrm{Eo}\right) = 1.0422\ 10^{-3}\mathrm{Eo}^3 - 1.59\ 10^{-2}\mathrm{Eo}^2 \\ -2.04\ 10^{-2}\mathrm{Eo} + 0.474 \tag{12}$$

Wall lubrication force

In the two-fluid Eulerian modeling approach, the near wall effects are captured by the wall lubrication forces $\mathbf{F}_{WL}$. These forces are accounted for repulsive effects pushing the bubbles away from the wall, referred usually as wall peaking effects. The mathematical expression of $\mathbf{F}_{WL}$ is given by:

$$\mathbf{F}_{WL} = -C_{WL}\alpha_v\rho_v\|\mathbf{U}_r - \left(\mathbf{U}_r.\mathbf{n}_w\right)\mathbf{n}_w\|^2\mathbf{n}_w \tag{13}$$

where $\mathbf{U}_r$ is the relative velocity, $\mathbf{n}_w$ is the unity vector normal to the wall and C_{WL} is the force model coefficient, calculated based on the models of Antal et al. (1991) and Franc (2005). The Antal et al. model is given by:

$$C_{WL} = \max\left(0, \frac{C_{w,1}}{d_b} + \frac{C_{w,2}}{y_w}\right) \tag{14}$$

where the parameters C_{w1} and C_{w2} are set to -0.01 and 0.05 respectively (Frank, 2005), d_b is the bubble diameter and y_w is the bubble center distance to the nearest wall. The Frank model is written as:

$$C_{WL} = C_{w,3}(\mathrm{Eo})\max\left(0, \frac{1 - \frac{y_w}{C_{w,c}d_b}}{y_w C_{w,d}\left(\frac{y_w}{C_{w,c}d_b}\right)^{p-1}}\right) \tag{15}$$

where $C_{w,3}(\mathrm{Eo})$ is given as following:

$$C_{w,3} = \begin{cases} \exp\left(-0.933\mathrm{Eo} + 0.179\right) & \text{if } \mathrm{Eo} \leq 5 \\ 5.99\ 10^{-3}\mathrm{Eo} - 0.0187 & \text{if } 6 < \mathrm{Eo} \leq 33 \\ 0.0179 & \text{else.} \end{cases} \tag{16}$$

with $C_{w,c}$, $C_{w,d}$ and p are a cut-off coefficient, a damping coefficient and a power, set to 10, 6.8 and 1.7, respectively (Frank, 2005).

Turbulent dispersion force

The turbulent dispersion force $\mathbf{F}_{TD}$ originates from the dispersions of bubbles, caused by fluctuations and eddies of the turbulent flow. The Lopez de Bertodano (1991) and the Burns et al. (2004) models are compared in this work. The Lopez de Bertodano model is mathematically expressed as:

$$\mathbf{F}_{TD} = C_{TD}\rho_l k_l \nabla \alpha_v \tag{17}$$

while the Burns et al. model is given by:

$$\mathbf{F}_{TD} = \frac{3}{4} C_{TD} \mathrm{Re}_b \frac{\nu_v \nu_v^t}{\sigma d_b^2} \rho_v \left(1 + \frac{\alpha_l}{\alpha_v}\right) \qquad (18)$$

where C_{TD} is the turbulent dispersion force coefficient, set to 0.4 and 1.0 for comparison, and ν_v and ν_v^t are the vapor phase kinematic and turbulent kinematic viscosities.

2.3 Boiling model

The boiling model adopted in this work follows the version of the very well known RPI (Rensselaer Polytechnic Institute) model proposed by Kurul and Podowski (1990). This model proposes to decompose the total applied heat flux q_w'' at the wall into three components; convective heat flux $q_{w,c}''$ that is responsible for heating the liquid near the surface when there are no bubbles in contact with it, evaporative heat flux $q_{w,e}''$ that is accounted for the liquid evaporation to create bubbles, and quenching heat flux $q_{w,q}''$ that represents the transient conduction while the liquid is filling the previous occupied volumes by departing bubbles. This decomposition is written as:

$$q_w'' = q_{w,c}'' + q_{w,e}'' + q_{w,q}'' \qquad (19)$$

The mathematical expressions of each heat flux contribution can be found in (Kurul and Podowski, 1990). The RPI model needs to be fed with closure equations that are applied at the heated surface, as the active nucleation site density N_a, the bubble departure diameter d_{dep} and the bubble departure frequency f_{dep}. These latter boiling closures are considerably affected by the heated surface roughness. It can increase the number of the nucleation sites, and controls the bubble shape and release rate. However, despite its major effect on bubble formation rates, the wall roughness will not be considered in this work, and the basic boiling closure equations neglecting this parameter are adopted here. The reformulated model of Lemmert and Chawla (1977) by Egorov and Menter (2004) is used to model here the active nucleation site density. It is given by:

$$N_a = N_{a,ref} \left(\frac{\Delta T_{sup}}{\Delta T_{ref}}\right)^{1.805} \qquad (20)$$

where the empirical coefficient $N_{a,ref}$ is set to $9.922 \ 10^5$ sites/m^2 following the recommendations of Egorov and Menter (2004), ΔT_{sup} is the wall superheat that is equal to the difference between the wall temperature and the liquid saturation temperature, and ΔT_{ref} is a reference temperature difference, set to 10 K. For the bubble departure diameter calculations, a modified version of the model of Tolubinski and Kostanchuk (1970) which is given as following is adopted:

$$d_{dep} = d_{ref} \exp\left(\frac{\Delta T_{sub,y^+=250}}{\Delta T_{ref}}\right) \qquad (21)$$

where d_{ref} and ΔT_{ref} are reference bubble departure diameter and temperature difference, set to $0.6 \ 10^{-3}$ m and 45

K, respectively, following the recommendations of Tolubinski and Kostanchuk (1970), and $\Delta T_{sub,y^+=250}$ is a modified subcooling temperature, calculated based on the liquid temperature at a distance from the wall based on the wall function at $y^+ = 250$. The last closure equation for the boiling model is the bubble departure frequency model, calculated based on the Cole (1960) model as:

$$f_{dep} = \sqrt{\frac{4g\left(\rho_l - \rho_v\right)}{3 d_{dep} \rho_l}} \qquad (22)$$

The last quantities calculated by the boiling model are the evaporation and the condensation rates Γ_{lv} and Γ_{vl}, that are needed to solve the transport equations. They are given by:

$$\Gamma_{l,v} = \frac{A_{w,b,e}^f}{6} \rho_v d_{dep} f_{dep} \qquad (23)$$

$$\Gamma_{v,l} = \frac{h_c \left(T_{sat}(p) - T_l\right)}{h_{lg}} A_s \qquad (24)$$

where $A_{w,b,e}^f$ is an area fraction of the heated surface not affected by bubbles and h_c is a condensation heat transfer coefficient calculated based on the correlation of Ranz and Marshall (1952).

3 Computational Domain and Solution Procedure

The computational domain is illustrated through Figure 1. It consists of 2D axisymmetric concentric pipe, having inner and outer diameters r_{in} and r_{out} equal to 7.89 mm and 19.01 mm, respectively. The concentric pipe is heated only from the inner upper side, by applying a constant uniform heat flux, allowing an adiabatic section for the flow development. Refrigerant R-113 flows downstream the adiabatic section of 0.91 m and then in 2.75 m heated section. The total flow length is 3.66 m. The computational domain is meshed using 20496 hexahedron elements, corresponding to 56 elements in the radial direction needed
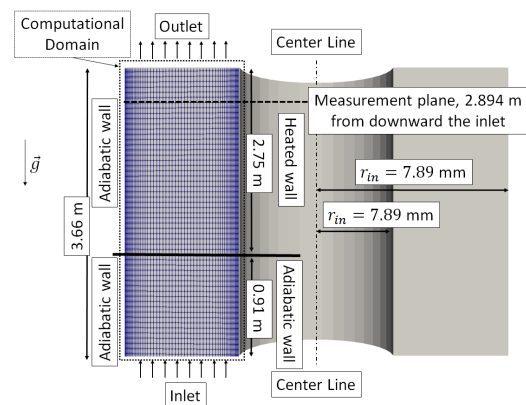

**Figure 1.** Computational domain and mesh.

for the grid convergence and 366 element in the flow direction. Near each wall, 20% of the radial length is meshed with refined elements to represent the viscous layers. The refined elements represent 30% of the radial elements with a expansion ratio of 0.25 at each wall normal direction.

At the pipe inlet, the pressure is set to 2.69 bar, corresponding to a saturation temperature of 80.5°C for R-113. Constant velocity and temperature are applied, set to 0.522 m/s and 50.2°C, respectively. This corresponds to an inlet Reynolds number and subcooling equal to 34450 and 30.3°C. A non-slip boundary condition is used for both phases velocities at the channel walls, and a constant uniform heat flux is applied only at the inner heated wall of length 2.75 m, set to 116000 W/m^2. Only in this part of the wall, the boiling closure equations are applied. At the channel outlet, a pressure gradient boundary condition is applied. The refrigerant R-113 thermophysical properties are calculated at the inlet based on the corresponding temperature and pressure. However, the developed CFD model allows the calculation of the local saturation temperature based on the corresponding predicted local pressure. For this purpose, an interpolation is performed on R-113 saturation tables.

The open-source CFD platform OpenFOAM is used to solve the transport equations with respect to the previous described computational domain and operating conditions. A Finite-Volume discretization technique is employed, where the spatial derivatives including the void fraction are discretized based on Van Leer scheme, the gradient and divergence terms are based on a Gauss upwind schemes, while the Laplacians are based on Gauss linear schemes. A semi-explicit temporal discretization is adopted for the temporal derivatives, in order to accelerate the convergence of the simulations. The discretized equations are solved based on the Geometric-Algebraic Multi-Grid (GAMG) solver for the pressure, and iterative symmetric Gauss-Seidel smooth solver is used for the rest of the variables. The PIMPLE algorithm is employed to solve the pressure-velocity coupling.

4 Results and discussions

The void fraction, liquid temperature, liquid and vapor magnitude velocity fields predicted by the CFD calculations are presented in Figure 2. Despite that the simulations were carried out in a 2D axisymmetric geometry, having a total flow length of 3.66 m, the fields were extruded rotationally and re-scaled for a better comprehensible representation. These results are obtained with calculations using Schiller and Naumann (1935) drag model, a constant virtual mass coefficient model equal to 0.5, Antal et al. (1991) wall lubrication model and Burns et al. (2004) turbulent dispersion model. These interfacial forces are adopted for the fields representation since they give the best predictions with comparison to the experimental data of Roy et al. (2001) as it will be shown later.
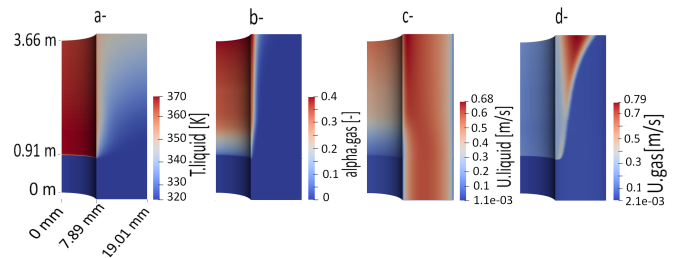
Figure 2. CFD fields spatial distributions: a- Liquid temperature, b- Void fraction, c- Liquid velocity and d- Vapor velocity.

The working fluid R-113 enters the concentric pipe at a uniform velocity distribution, with a mean equal to 0.522 m/s, and an inlet temperature equal to 50.2°C, representing 30.3°C below the saturation temperature. Along the adiabatic section, from the inlet downward to 0.91 m, where heating starts, the void fraction showed through Figure 2(b) maintains a zero value, meaning that there is no bubble formation and no phase change, corresponding to a constant liquid temperature equal to the inlet temperature, as presented by Figure 2(a). As can be deduced from Figure 2(c), the liquid velocity in the adiabatic flow section follows a parabolic profile, where it is zero in the viscous boundary layer at the wall, where the non-slip boundary condition is applied. In this section, the flow is being developed, with an average Reynolds number at 34450, referring to a turbulent flow. In the heated flow section, the void fraction starts having values higher than zero, corresponding to bubbles nucleation at the heated surface. Nucleation starts when the local liquid temperature in the heated surface vicinity reaches the saturation temperature, corresponding to the local system pressure. The void fraction at the heated surface increases with increasing heated length, up to a maximum value of 40% near the outlet. However, it decreases with increasing radial distance from the heated surface, to become zero far away from it. This is the typical subcooled nucleate boiling heat transfer regime, where the bubbles are formed at the heated surface, giving the highest radial void fraction. This latter will decrease when increasing the radial distance from the heated surface, due to the absence of evaporation far from the wall, and bubbles are subject to condensation, since the liquid is still subcooled and the surrounding bubbles temperature gradually drops below the saturation. This also can be proved by the liquid temperature distribution presented in Figure 2(a). When entering the heated flow section at a temperature equal to the inlet temperature, the liquid near the heated surface starts to heat up until it reaches its saturation temperature, at which evaporation at the wall occurs. Far from the heated surface in the radial direction, the liquid is heated, but it does not reach its saturation temperature and is still subcooled. The bubbles which left their nucleation sites will condense in this flow region. A special behavior of the liquid phase velocity at the viscous boundary layer near the heated surface is observed. The liquid velocity reaches its maximum
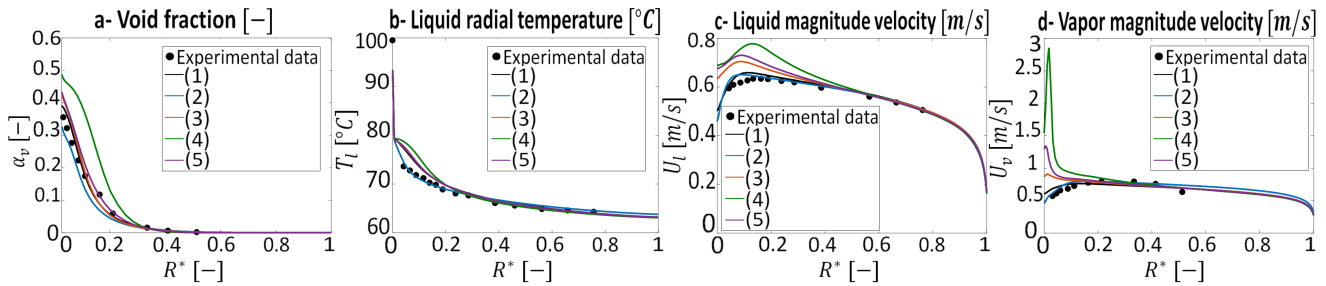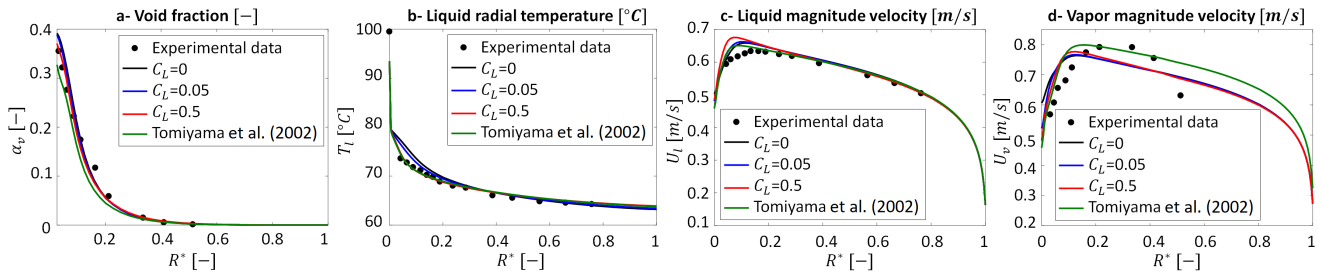
Figure 3. Interfacial forces effect on the radial boiling fields. Experimental data of Roy et al. (2001). a- Void fraction prediction, b- Liquid radial temperature, c- Liquid magnitude velocity and d- Vapor magnitude velocity. Legend: (1) Schiller and Naumann (1935) drag force model, Antal et al. (1991) wall lubrication force model and Burns et al. (2004) turbulent dispersion force model with $C_{TD} = 1$. (2) Tomiyama et al. (1998) drag model, Frank (2005) wall lubrication force model and Burns et al. (2004) turbulent dispersion force model with $C_{TD} = 1$. (3) Schiller and Naumann (1935) drag force model, neglected wall lubrication force and Burns et al. (2004) turbulent dispersion force model with $C_{TD} = 1$. (4) Schiller and Naumann (1935) drag force model, neglected wall lubrication force and Lopez de Bertodano (1991) turbulent dispersion model with $C_{TD} = 0.4$. (5) Schiller and Naumann (1935) drag force model, neglected wall lubrication force and Lopez de Bertodano (1991) turbulent dispersion force model with $C_{TD} = 1$.

near the wall, under the effect of the bubble nucleation and motion. This is a feature of boiling flows, where the liquid velocity which approaches zero near the adiabatic walls becomes increasingly higher at the heated section of the wall, away from the non-slip boundary condition. This yields increased heat transfer coefficients hence enhancing the heat transfer from the heated wall to the bulk liquid. At the external adiabatic wall, in both flow regions, adiabatic and heated, the liquid velocity keeps the parabolic profile, where it is very low in the viscous boundary layer. Concerning the vapor phase velocity, Figure 2(c) shows that the vapor velocity increases with increasing void fraction and the two-phase layer thickness.

The CFD predictions of the radial profiles of the quantities of concern, i.e., void fraction, liquid temperature, liquid and vapor velocities, following the measurement section used by Roy et al. (2001), located at 2.894 m downward the concentric pipe inlet are presented through Figure 3. The dimensionless distance R^* is calculated as $(r - r_{in})/r_{out}$. These predicted fields are compared against the available experimental data of Roy et al. (2001). In order to quantify the interfacial forces effect on the CFD simulation results, different interfacial forces were taken into account or neglected, employing the most relevant force models for boiling. In Figure 3, the studied forces are the drag force, the wall lubrication force and the turbulent dispersion force. The drag force is taken into account in all the simulations, where the models of Schiller and Naumann (1935) and Tomiyama et al. (1998) are compared, while the turbulent dispersion force is modeled based on Burns et al. (2004) and Lopez de Bertodano (1991). However, the wall lubrication force is taken into account in some simulations only and neglected in others, in order to quantify its effect on boiling. The models of Antal et al. (1991) and Frank (2005) are tested. The results presented by Figure 3 are obtained for a constant added mass coefficient model set to 0.5, and a neglected lift force. As it can be seen in Figure 3(a), a successful prediction of the void fraction with comparison to

the experimental data is obtained with the drag model of Schiller and Naumann (1935), the wall lubrication force of Antal et al. (1991) and the turbulent dispersion force of Burns et al. (2004), with a turbulent dispersion coefficient $C_{TD} = 1.0$. These interfacial forces models give also fair predictions of the liquid and vapor phase velocities, presented by Figures 3(c) and 3(d), respectively. However, using these models will yield to an over-estimation of the liquid temperature in the two-phase boundary layer near the heated surface, as it can be seen in Figure 3(b). The successful predictions of the liquid and vapor phase velocities and the liquid phase temperature are obtained by the predictions associated to the drag model of Tomiyama et al. (1998), the wall lubrication model of Frank (2005) and the turbulent dispersion model of Burn et al. (2004), with $C_{TD} = 1.0$. Nevertheless, these models yield to a small underestimation of the void fraction in the vicinity of the heated surface. Then, neglecting the wall lubrication force will affect considerably the predictions accuracy. Based on the comparison against the available experimental data, the predictions associated with neglecting the wall lubrication force overestimate all the considered fields. The turbulent dispersion model of Lopez de Bertodano (1991) gives always high over-prediction, and the highest one is associated with a lower turbulent dispersion force coefficient of C_{TD}. This drag model with a lower C_{TD} is giving also the highest over-prediction of the liquid radial temperature and velocity. However, a failed prediction of the vapor phase velocity is obtained when the Lopez de Bertodano (1991) turbulent dispersion model is employed, regardless of the value of C_{TD}, especially in the two-phase boundary layer near the heated surface. Despite the overestimations of the Burns et al. (2004) turbulent dispersion model when neglecting the wall lubrication force, the predictions are fair for the void fraction and the liquid radial temperature. Nonetheless, for the liquid and the vapor phase velocities, the predictions associated with the latter mentioned models are failed, especially at the heated surface, where overly high velocities are obtained. The

Figure 4. Lift force effect on the radial boiling fields. Experimental data of Roy et al. (2001) a- Void fraction prediction, b- Liquid temperature, c- Liquid velocity, d- Vapor velocity.

two-phase boundary layer thickness, where bubbles are present and subject to nucleation and condensation, is very well predicted using any interfacial force model, as it can be deduced from the estimated void fraction comparison with the experimental data through Figure 3(a). Hence, when neglecting the lift force, whose effect on boiling will be discussed in the next paragraph, and setting a constant added mass coefficient equal to 0.5 as well as neglecting the wall lubrication force will yield an over-estimation of all the considered fields, and this overestimation is amplified when a low turbulent dispersion model is used. The models yielding the best void fraction prediction will over-predict the other fields considered, and the successful predictions of the liquid radial temperature, liquid and vapor velocities will give a small under-prediction of the void fraction.

The lift force effect on the CFD predicted fields is quantified in Figure 4. Two different lift models are compared here, a constant lift coefficient model, set to 0.05 and 0.5, and the lift coefficient model of Tomiyama et al. (2002). When a constant lift coefficient model is employed, then the Schiller and Naumann (1935) drag force model, the Antal et al. (1991) wall lubrication force model and the Burns et al. (2004) turbulent dispersion force model are used. For the Tomiyama et al. (2002) lift model, Tomiyama et al. (1998) drag model, the Frank (2005) wall lubrication force model and Burns et al. (2004) turbulent dispersion force are employed. In all the performed simulations of Figure 4, the added mass force was taken into account via a constant value set to 0.5. The comparison with the experimental data shows that deviations between the tested models are minor. The Tomiyama et al. (2002) lift model gives the best predictions for the liquid radial temperature, liquid velocity and vapor velocity, with comparison to the experimental data, presented respectively through Figures 4(b), 4(c) and 4(d). However, a small underestimation is observed for the void fraction illustrated in Figure 4(a). This trend is observed in the results of Figure 3(a) when the Tomiyama et al. (1998) drag model, the Frank (2005) wall lubrication model and the Burns et al. (2004) turbulent dispersion model are used. The authors believe that this underestimation is caused by the interfacial models, except with the Tomiyama et al. (2002) lift model, and that its effect on the CFD calculations is mi-

nor. Despite the small differences obtained when different constant lift coefficients C_L are used, the best prediction is obtained when $C_L = 0.5$ for void fraction. However, the corresponding liquid velocity is very high near the heated surface, and this will yield to over-predictions of the heat transfer. Comparable results are observed when performing the calculations with $C_L = 0.05$ and neglecting the lift ($C_L = 0$). Hence, despite its importance in two-phase bubbly flows, the lift force has a minor effect when phase change is considered. Hence, this force can be neglected when modeling boiling flows.

5 Conclusions

In this work, 2D axisymmetric CFD simulations were carried out to model the subcooled nucleate boiling flow of refrigerant R-113 flowing upward a narrow concentric pipe. The void fraction spacial distribution showed that boiling starts at the heated section, and the highest values are obtained near the heated surface where nucleation occurs, corresponding to the local saturation temperature. The temperature field showed that the saturation temperature is reached near the heated surface, and that the liquid stills subcooled away from the heated surface, which allows the bubbles to leave the heated surface and condense. This will yield to decreased values of the void fraction, to reach zero at the end of the two-phase layer. The velocity field proved that higher values are obtained at the heated surface, under the effect of the bubbles nucleation and motion, and this is one of the most important boiling flows features, leading to an increased values of the heat transfer coefficient, that was very well predicted by the current CFD simulations.

A sensitivity analysis of the acting interfacial forces models was performed by comparing the obtained CFD predictions with the experimental data of Roy et al. (2001). It was shown that all the interfacial forces, except the lift, have a considerable effect on the predicted fields. Neglecting the wall lubrication force will yield to increased values of the void fraction, liquid radial temperature near the heated surface and liquid phase velocity. This will lead to failed predictions of the vapor phase velocity. It was shown also that the models giving the best void fraction predictions will give small overestimations of the other fields, and models giving successful predic-

tions of the liquid radial temperature and the velocities of the liquid and vapor phases will slightly under-predict the corresponding void fraction. It was proven that the lift force has a minor effect on the predicted fields, and it can be neglected in CFD simulations of subcooled boiling flows without any major concern.

The present work provides a knowledge foundation for performing CFD simulations of subcooled nucleate boiling flows of refrigerants. It provides guidance on the corresponding interfacial forces that should be taken into account, with model recommendations provided for better accuracy with respect to each important boiling regions. Future improvement of this work can be recapitalized in using IATE model to predict the bubbles bulk diameter instead of a constant model, as adopted in this work. Also, the interfacial forces analysis should be complemented by turbulence model quantification.

Acknowledgments

References

H. Anglart and O. Nylund. CFD application to prediction of void distribution in two-phase bubbly flows in rod bundles. *Nuclear Engineering and Design*, 163(1):81 – 98, 1996. doi:10.1016/0029-5493(95)01160-9.

S. P. Antal, R. T. Lahey Jr, and J. E. Flaherty. Analysis of phase distribution in fully developed laminar bubbly two-phase flow. *International Journal of Multiphase Flow*, 17 (5):635 – 652, 1991. doi:10.1016/0301-9322(91)90029-3.

A. D. Burns, T. Frank, I. Hamill, and J. M. Shi. The Favre averaged drag model for turbulent dispersion in Eulerian multiphase flows. In *Proceedings - 5th International Conference on Multiphase Flow, ICMF'04*, Yokohama, Japan, May 2004.

R. Cole. A photographic study of pool boiling in the region of the critical heat flux. *AIChE Journal*, 6(4):533 – 538, 1960. doi:10.1002/aic.690060405.

Y. Egorov and F. Menter. Experimental implementation of the RPI wall boiling model in CFX-5.6. Technical report, ANSYS/TR-04-10, 2004.

T. Frank. Advances in computational fluid dynamics (CFD) of 3-dimensional gas-liquid multiphase flows. In *Proceedings - NAFEMS Seminar: "Simulation of Complex Flows (CFD)"*, Niedernhausen/Wiesbaden, Germany, April 2005.

M. Ishii, S. Kim, and J. Kelly. Development of interfacial area transport equation. *Nuclear Engineering and Technology*, 37 (6):525–536, 2005.

B. Končar and E. Krepper. CFD simulation of convective flow boiling of refrigerant in a vertical annulus. *Nuclear Engineering and Design*, 238(3):693 – 706, 2008. doi:10.1016/j.nucengdes.2007.02.035.

B. Končar, I. Kljenak, and B. Mavko. Modelling of local two-phase flow parameters in upward subcooled flow boiling at low pressure. *International Journal of Heat and Mass Transfer*, 47(6):1499 – 1513, 2004. doi:10.1016/j.ijheatmasstransfer.2003.09.021.

N. Kurul and M. Z.. Podowski. Multidimensional effects in forced convection subcooled boiling. In *Proceedings - International Heat Transfer Conference Digital Library*. Begel House Inc., 1990.

J.C. Lai and B. Farouk. Numerical simulation of subcooled boiling and heat transfer in vertical ducts. *International Journal of Heat and Mass Transfer*, 36(6):1541 – 1551, 1993. doi:10.1016/S0017-9310(05)80064-X.

M. Lemmert and J. M. Chawla. Influence of flow velocity on surface boiling heat transfer coefficient. *Heat Transfer in Boiling*, pages 237–247, 1977.

M. Lopez de Bertodano. *Turbulent bubbly flow in a triangular duct*. PhD thesis, Rensselaer Polytechnic Institute, New York, USA, 1991.

E. Michta. Modeling of subcooled nucleate boiling with Open-FOAM. Master's thesis, KTH Royal Institute of Technology, Stockholm, Sweden, 2011.

A. Rabhi and R. Bel Fdhila. Evaluation and analysis of active nucleation site density models in boiling. In *Proceedings - The Second Pacific Rim Thermal Engineering Conference*, Maui, Hawaii, USA, December 2019.

W. E. Ranz and W. R. Marshall. Evaporation from drops. *Chemical Engineering Progress*, 48(3):141–146, 1952.

R. P. Roy, S. Kang, J. A. Zarate, and A. Laporta. Turbulent Subcooled Boiling Flow—Experiments and Simulations . *Journal of Heat Transfer*, 124(1):73–93, 2001. doi:10.1115/1.1418698.

Y. Sato, M. Sadatomi, and K. Sekoguchi. Momentum and heat transfer in two-phase bubble flow—II. A comparison between experimental data and theoretical calculations. *International Journal of Multiphase Flow*, 7(2):179 – 190, 1981. doi:10.1016/0301-9322(81)90004-5.

L. Schiller and Z. Naumann. A drag coefficient correlation. *Zeit. Ver. Deutsch. Ing.*, 77:318–320, 1935.

V. I. Tolubinsky and D. M. Kostanchuk. Vapour bubbles growth rate and heat transfer intensity at subcooled water boiling. In *Proceedings - International Heat Transfer Conference 4*, volume 23. Begel House Inc., 1970.

A. Tomiyama, I. Kataoka, I. Zun, and T. Sakaguchi. Drag coefficients of single bubbles under normal and micro gravity conditions. *JSME International Journal*, 41(2):472–479, 1998.

A. Tomiyama, G. P. Celata, S. Hosokawa, and S. Yoshida. Terminal velocity of single bubbles in surface tension force dominant regime. *International Journal of Multiphase Flow*, 28 (9):1497 – 1519, 2002. doi:10.1016/S0301-9322(02)00032-0.

N. Zuber. On the dispersed two-phase flow in the laminar flow regime. *Chemical Engineering Science*, 19(11):897 – 917, 1964. doi:10.1016/0009-2509(64)85067-3.

Study of agricultural waste gasification in an air-blown bubbling fluidized bed using a CPFD model

Rajan Jaiswal   Rajan K. Thapa   Britt M. E. Moldestad

Department of Process, Energy and Environmental Technology, University of South-Eastern Norway, Norway,
{rajan.jaiswal rajan.k.thapa, britt.moldestad}@usn.no

Abstract

Gasification using a fluidized bed is a promising technology to convert agricultural residues into product gases. In this work, the syngas production potential from agricultural waste (grass pellets) is studied using a Computational Fluid Dynamic (CPFD) model. The CPFD model is developed in a simulation software Barracuda virtual reactor and validated against the experimental data. Experiments are carried out in a 20 kW bubbling fluidized bed gasification reactor that operates with air as fluidizing gas. Grass pellets of size 5mm-30 mm in length and diameter of 5mm are used as the feed.

The CPFD model considers the hydrodynamics of the gas-solid phase and reaction kinetics involved. Influence of the static bed height, bed temperature, and air to fuel ratio on the product gas composition ($CH_4, CO, CO_2, and\ H_2$) and char conversion efficiency are investigated. Initial bed heights of 200 mm and 300 mm are used for the analysis. Biomass is fed at 2.46 kg/hr while the air supplied is varied to obtain the air to fuel ratio at 0.4, 0.6, 0.8, 1, and 1.2. The result shows that the increase in bed height has a significant effect on the reactor temperature but very small effect on the product gas composition and char conversion rate. An increase in bed temperature from 600℃ to 800℃ improves the gasifier performance in terms of maximum product gases yield and enhanced char conversion rate. Increase in the air to fuel ratio from 0.4-1.2 reduces the $CH_4, CO,\ and\ H_2$ fractions in the product gas and increases the CO_2 concentration. The results obtained from the CPFD model are in good agreement with the experimental results and literature data. Thus, the CPFD model developed in this work can be utilized to optimize the gasification reactor used in a lab and industrial scale.

Keywords: grass pellets, gasification, bubbling-fluidized bed, CPFD

1 Introduction

With the world's growing population, increase in solid wastes from households, agriculture and industries are inevitable. Strict laws and regulations are set up in the EU regions and other countries like USA and Japan to reuse and recycle the municipal solid waste (MSW).

Despite such strict measures three quarters of the MSW are still incinerated or landfilled. Converting such wastes into energy not only preserves the landfill space, but also facilitates the increasing energy demand. Thermochemical conversion technologies such as pyrolysis, combustion, and gasification enable efficient conversion of solid wastes to different energy forms such as heat and electricity. Gasification of solid wastes has several potential benefits over combustion specifically in terms of operating conditions and various reactor types suited for specific purposes. In the gasification process, the solid wastes are partially oxidized with limited amount of oxygen or steam that prevents combustion. The conversion of solid waste to product gases in a gasification process occurs in complex thermochemical routes. The first step is drying and devolatilization. The solid fuels are converted into volatile gases, char, and tar after pyrolysis and then the char is gasified with a gasifying agent (air or steam) at temperatures in the range of 600-900℃. The process can be autothermal or allothermal, depending on the gasifying agent and the type of reactors. When using air as a gasifying medium, the process is driven auto-thermally. The heat required for the chemical conversion of the fuel is supplied by partial oxidation reactions. The exothermic reactions are absent when steam is used as the gasifying agent, and thus an external heat source is needed. The advantages of using steam as gasifying medium is that a synthesis gas with a higher heating value $(15 - 20\ MJ/Nm^3)$ is produced. When using air, the caloric value of the synthesis gas is in the range of $4 - 8\ MJ/Nm^3$. The advantage of using air as fluidizing gas is that it is cheaper, and some gas turbines available in the market enable to use lower calorific value syngas to produce electricity.

The main gas components obtained from the gasification process are methane, hydrogen, carbon monoxide and carbon dioxide. The product gases can be useful for producing biofuels and chemicals or can be used to operate gas turbines and reciprocating engines. Gasification using fluidized bed has been considered as an flexible technology that can use a wide range of fuels. Fluidized beds provide major advantages such as uniform mixing and heat transfer (Jaiswal, 2018; Jaiswal, 2019), which enables efficient conversion of carbonaceous solid into product gases. The feedstock to the gasifier has to be cheaper and more flexible in order to make the process economically sustainable. Thus,

utilizing wastes from forest, agriculture and household as feed reduces the overall costs. Over the past few years, industrial companies and researchers around the world have assessed the benefits of converting solid waste into energy and valuable chemical products. Many researches have focused on investigating different parameters that influence the product gas composition and the gasifier efficiency.

Arena *et al.* performed gasification tests with different types of waste-derived fuel in a pilot scale bubbling fluidized bed using air as the gasifying agent and olivine and quartz sand as the bed material. The authors reported that gasification of polyolefin plastic wastes increases the hydrogen concentration by (20-30%) in the syngas. In a similar study with solid recovered fuel (SRF), the authors concluded that the SRF could be gasified to obtain a syngas of valuable quality for energy applications (Arena *et al.*, 2010).

Xiao *et al.* studied gasification of polypropylene plastic waste in a fluidized bed gasifier of diameter 100 mm and a height of 4.2 m. The authors studied the effect of equivalence ratio, bed height and fluidization velocity on the product yield distribution, gas compositions, and gas-heating value. They found that polypropylene with air as the gasifying agent, can produce a fuel gas with a calorific value in the range of 5.2-11.4 MJ/Nm³. (Xiao *et al.*, 2007)

Alvarez *et al.* studied hydrogen production from gasification of a mixture of plastic and wood sawdust using steam. Adding 20% of polyethylene to the biomass (sawdust), they found that the gas yield and the hydrogen concentration increased due to the synergetic effects between increased gas components. (Alvarez *et al.*, 2014)

Predicting performance of the fluidized bed gasifier with a simulation model is of great importance for optimum design of the reactor to achieve maximum efficiency. The CFD model can provide detail knowledge about the operational parameters to be selected. Several researches have been conducted in the fluidized bed gasification using simulation tools (Niu, 2013; Erkiaga, 2014; Shen, 2008; Chen, 2016). However, most of the studies available in literature, either lack experimental validation of the simulation model or the model does not account for the particle size distribution of the bed material and fuel particles. Particle size distribution is an important parameter that largely influences the heat and mass transfer within the bed (Grace, 1991; Beetstra, 2009) thus affect reactor performance.

In this work, a Computational Particle Fluid Dynamic (CPFD) software Barracuda VR is used. It is a commercial simulation software specially designed for the particle-fluid system. Unlike other CFD software, it is possible to define computational particles similar to the experiments with a size distribution in Barracuda. Many studies have been devoted to gain more insight into the gasification of biomass in the past using Barracuda VR (Jaiswal, 2020; Bandara, 2020; Thapa, 2016). However, limited researches are available in the field of gasification of solid wastes using a CPFD tool. In this work, the syngas production potential from agricultural waste (grass pellets) is studied using a CPFD model. The CPFD model is validated against experimental data. The CPFD model considers the hydrodynamics of the gas-solid phase and reaction kinetics involved. The influence of the static bed height, the bed temperature and the air to fuel ratio on the product gases $(CH_4, CO, CO_2, and\ H_2)$ and char conversion efficiency are investigated.

2 Experiment

The experimental set up used in this work consists of a 20kW reactor made up of stainless steel. The reactor is 1 m in height and the internal diameter is 0.1 m. Three electrical heaters are mounted on the reactor wall and heat up the reactor during operation. An air preheater is used to heat the compressed air (fluidizing gas) before it is passed into the reactor. Thermocouples and pressure sensors are placed at different heights to measure the temperature and pressure variation during the operation of the reactor. The schematic diagram of the reactor is shown in Figure 1. The biomass feedstock is supplied into the reactor by the means of two screw conveyors. A cold screw conveyer transports the biomass from the silo to the hot screw conveyer at a specified rate. The hot screw conveyor runs continuously and injects biomass into the reactor. The bed material is added to the reactor with a funnel type opening (2) attached to the wall of the reactor. A sampling line is attached at the outlet of the reactor. The gas samples are collected at certain intervals, and are analyzed in an offline gas chromatography (GC). The GC uses helium and nitrogen as the carrier gas. The hydrogen concentration is measured using nitrogen as the carrier gas, while methane, carbon monoxide, and carbon dioxide and nitrogen concentration are measured using helium as the carrier gas.

Experiments were carried out using grass pellets of 5 mm diameter and 5-30 mm in length. Sand is used as the bed material and air as the fluidizing gas.
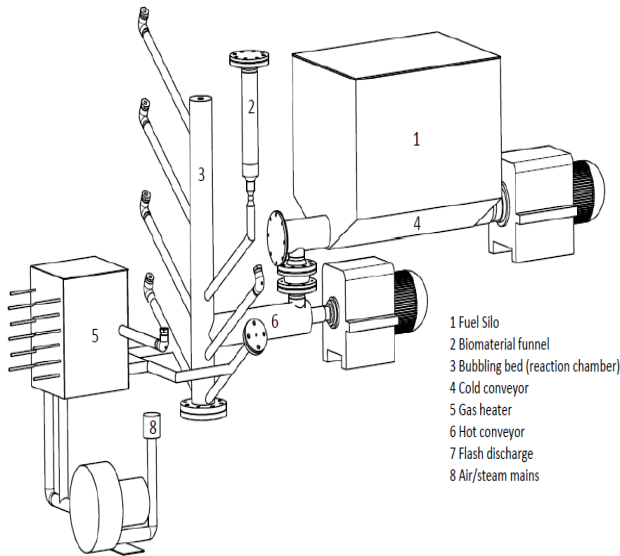
Figure 1. Experimental setup.

3 Result and Discussion

3.1 CPFD model

The developed CPFD model uses three-dimensional multiphase particle-in-cell approach for the CFD simulation of gas-particle flows. The chemistry module available in Barracuda enables to define the reactions and reaction rates involved in the gasification process. Therefore, the CPFD model simulates the thermal and chemical kinetics at the particle level providing more realistic modeling compared to other simulation software. The chemical reactions and corresponding rate kinetics used in the CPFD model are listed in Table 1.

Figure 2. Transient data points, flux plane, flow and pressure boundary conditions and feed position (From left)

A CAD geometry in STL format, drawn in SolidWorks was imported, and total cells of 7128 were created using the default grid setting available in Barracuda. The properties of fuel particles (grass pellets), bed material (sand), and fluidizing gas (air) are

defined as that of the experimental data. The pressure and flow boundary conditions (of air and fuel particles), transient data points, and the flux plan defined in the CPFD model are shown in Figure 2. Grass pellets (the fuel particles) of sizes 5-30cm in length and 5mm in diameter are used. It has a bulk density of 743 $kg/m3$ and a lower heating value of 16.7 MJ/kg. Sand particles of density 2600 $kg/m3$ and mean diameter $345\mu m$ are used as the bed material. For the simulations, the drag model proposed by Wen Yu is used. The drag model of the particle calculates the force acting on the particle by the flow of fluid around it. The simulated product gas compositions are compared with the experimental data for validation of the model. The comparison is shown in Figure 3. The result shows that the model can predict the product gas compositions close to the experimental data.

Table 1. Chemical reactions and reaction rate (Xie, 2016; Bates, 2017; Soli, 2016).

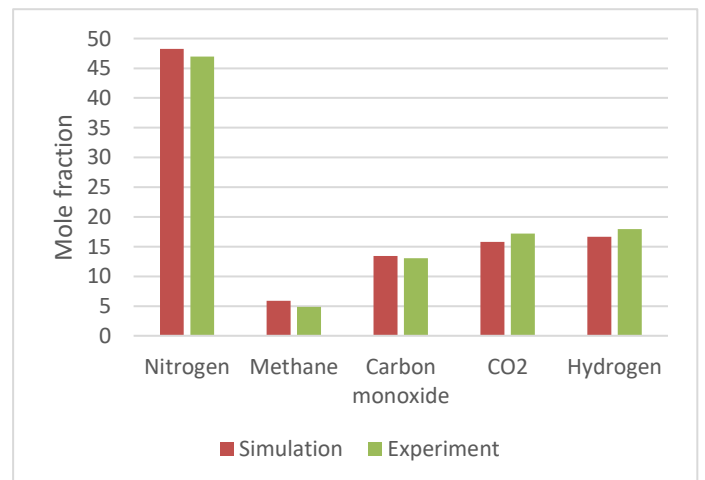| Chemical reactions	Kinetics
Char partial combustion R1:$2C + O_2 \leftrightarrow 2CO$	r = 4.34×10^{10} $m_s\theta_f$ $\exp\left(\frac{-13590}{T}\right)[O_2]$
CO oxidation R2:$CO + 0.5O_2 \leftrightarrow CO_2$	r = 5.62×10^{12} $\exp\left(\frac{-16000}{T}\right)[CO][O_2]^{0.5}$
H$_2$ oxidation R3:$H_2 + 0.5O_2 \leftrightarrow H_2O$	r = 5.69×10^{11} $\exp\left(\frac{-17610}{T}\right)[H_2][O_2]^{0.5}$
CH$_4$ oxidation R4:$CH_4 + 2O_2 \leftrightarrow CO_2 +$ $2H_2O$	r = 5.0118×10^{11} T^{-1} $\exp\left(\frac{-24357}{T}\right)[CH_4][O_2]$
Water gas shift reaction R5:$CO + H_2O \leftrightarrow CO_2 + H_2$	r = 7.68×10^{10} T $\exp\left(\frac{-36640}{T}\right)[CO]^{0.5}[H_2O]$
Methane reforming R6:$CH_4 + H_2O \leftrightarrow CO +$ $3H_2$	r=$3.00\times105\exp\left(\frac{-15042}{T}\right)[$ $CH_4][H_2O]$

Figure 3. Product gas compositions obtained from the CPFD model and experiments (model validation).

3.2 Effect of static bed height

Change in static bed height in a bubbling fluidized bed gasifier effects the bed hydrodynamic behavior (Jaiswal, 2018), the residence time of the fuel particles and the temperature of the bed. In this section, influence of the static bed height on the product gas composition and the reactor performance are discussed.

Figure 4 shows the product gas composition at static bed height of 300 mm and 200 mm. For both the cases, the air to fuel (A/F) is 0.6 and the reactor is operated at 700°C. The result shows that with increase in bed height from 200 mm to 300 mm at the same operating conditions the methane, carbon monoxide and hydrogen concentration in the product gas increases while the carbon dioxide concentration remains the same. The total heat content of the system increased with increased bed height. When the grass pellets are injected into the reactor at fluidizing conditions, the fuel particles first sink downwards in the bed and then tends to move upward due to densities differences (Jaiswal, 2019; Agu, 2019). The increase in bed height allows the fuel particles to interact with the heated bed material and the fluidizing gas for a longer time when the contact area is increased. Thus, the residence time of the fuel particles is increased, which favored the breakdown of heavy hydrocarbons, tar and char when exposed to high temperature. A similar trend is found in the experimental work performed by (Xiao *et al.*, 2007). They tested three initial bed heights and confirmed that for an equivalence ratio maximum gas yield is obtained at a specific bed height.
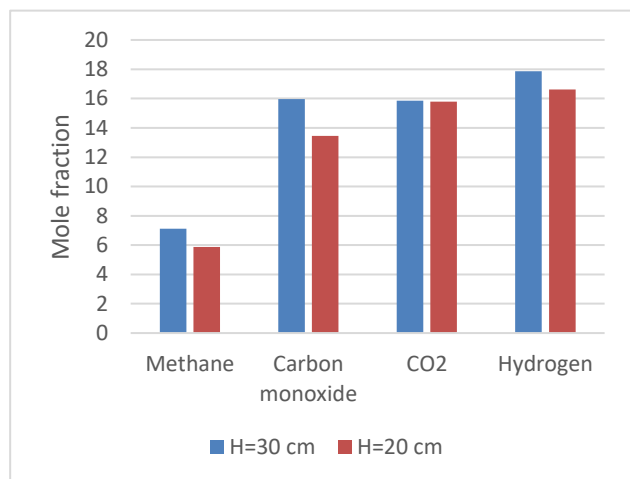
Figure 4. Change in the product gas composition with initial bed height.

3.3 Effect of equivalence ratio

Equivalence ratio is defined as the ratio of air to fuel supplied divided by the air to fuel ratio required for stoichiometric combustion, and is one of the important parameters in gasification of biomass with air. The equivalence ratio influences the product gas quality and quantity. Increase in equivalence ratio increases the amount of oxidant in the gasifier, and thus it influences the conversion of char and tar into the product gases depending on the temperature of the reactor.

In this work, the fuel supply (grass pellets) to the reactor is kept constant, while increasing the air flowrate and thereby increasing the equivalence ratio. Figure 5 shows the change in product gas composition at increased air to fuel ratio from 0.4 to 1.2. The concentration of methane, carbon monoxide and hydrogen is decreased and carbon dioxide concentration is increased with increase in the A/F ratio. This is because more oxygen is added to the system that accelerates the carbon monoxide, hydrogen and methane oxidation reaction routes (R2-R4). In addition, the elevated nitrogen into the reactor with increased air to fuel ratios dilutes the product gas composition and reduces its heating value.
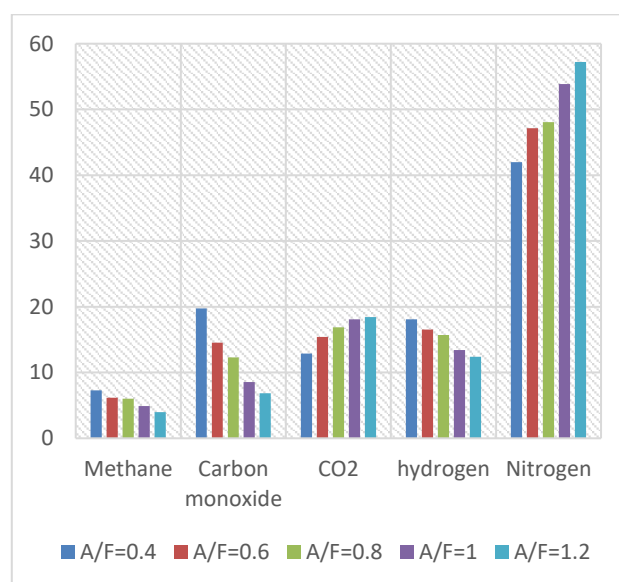
Figure 5. Product gas compositions at different air to fuel ratio.

3.4 Effect of temperature

Temperature in the gasifier determines the rate of chemical reactions involved in the gasification of biomass. The rate at which biomass is converted into volatile components, char, tar and finally into the product gases is influenced by the temperature in the bed and freeboard zone of the reactor. Lower temperature can end up with too much tar in the product gas and slow conversion rate of the char. While too high temperature may allow the ash to melt and stick together the bed materials, forming agglomerates and defluidization of the bubbling fluidized bed. Thus, the reactor must be operated at an optimum temperature so that the fuel particles are converted into better quality product gases with maximum gasifier efficiency.
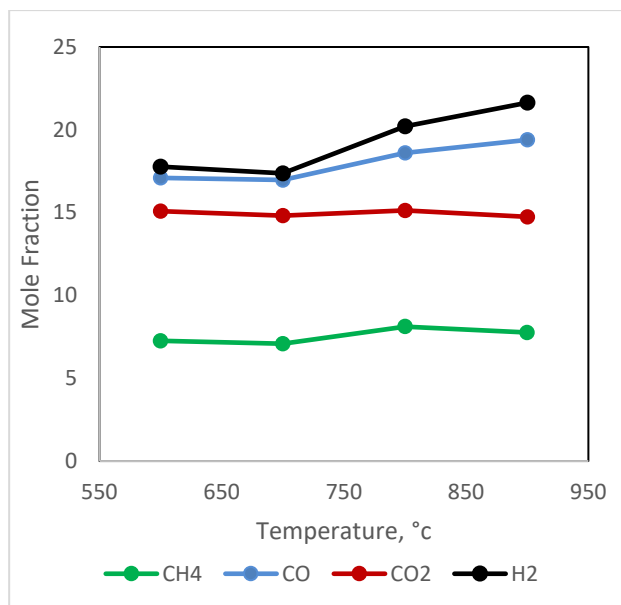
Figure 6. Product gas compositions at different reactor temperatures.

Four cases with temperatures 600°C, 700°C, 800°C, and 900°C are tested at air flowrate of 2 kg/hr and grass pellets feed rate at 2.46kg/hr. The initial bed height is 200mm for all the cases. Figure 6 depicts the product gas composition obtained at different reactor temperatures. With change in temperature from 600°C to 900°C, the concentration of hydrogen and carbon monoxide increased significantly and the methane fraction increased slightly. The reason is that the reaction routes (shown in Table 1), char partial oxidation (R1), water-gas-shift reaction (R5) and methane reforming (R6), are enhanced with the increase in temperature. The fraction of hydrogen and carbon monoxide increased from 17% to 21% and 19% respectively. Increase in quantity of higher calorific value gases enhance the lower heating value (LHV) of the product gas. The LHV and cold gas efficiency (CGE) of the product gas with change in temperature is shown in Figure 7. LHV of the product gas increases from 6.8 MJ/Nm^3 to 7.5 MJ/Nm^3. Since the CGE of the product gas is directly related to LHV of the gas (shown in equation 1), the CGE of product gas increases with increase in LHV of the gas. The CGE of the product gas increased from 58% to 66.6%.

$$CGE(\%) = \frac{LHV\ of\ the\ syngas * syngas\ flow\ rate * 100}{LHV\ of\ the\ fed\ grass\ pellets * grass\ pellets\ feed\ rate} \qquad (1)$$

$$LHV\left(\frac{MJ}{Nm^3}\right) = \frac{CO*126.36 + H_2*107.98 + CH_4*358.18}{1000} \qquad (2)$$
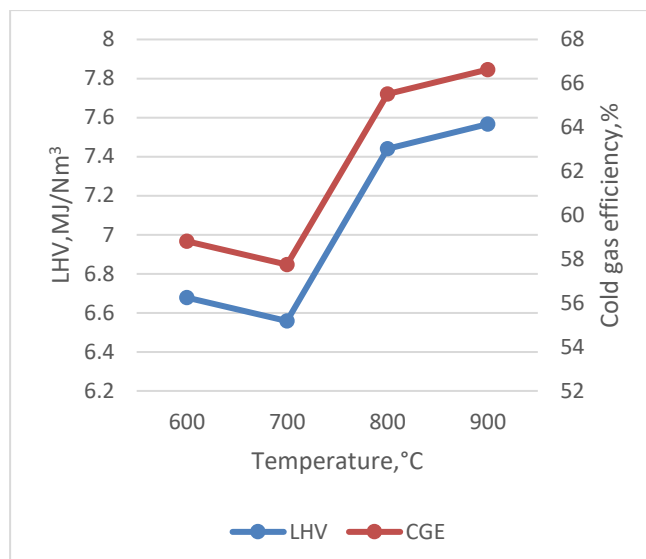
Figure 7. Lower heating value and cold gas efficiency of the product gas at different reactor temperatures.

4   Conclusion

The main objective of this work was to investigate the syngas production potential from gasification of agricultural wastes (grass pellets) using a commercial CPFD simulation software Barracuda VR. Experimental data were used to validate the CPFD model. Experiments were conducted on a 20 kW bubbling fluidized bed gasifier with air as a fluidizing gas and sand as the bed material. Grass pellets were fed to the reactor at 2.46 kg/hr. The samples were collected and measured in an offline GC. The product gas compositions obtained from the simulation results and the experimental data were compared for the model validation.

The influence of initial bed height, temperature and air to fuel ratio on the product gas compositions and gasifier performance were investigated. The result shows that with increase in initial bed height from 200 mm to 300mm there is a slight increment in the product gas yield and a significant increment in the reactor temperature. Different air to fuel ratios, 0.4, 0.6, 0.8, 1, 1.2, were simulated and the result shows that the quality of the product gas decreases with increase in the air to fuel ratio from 0.4 to 1.2. More carbon dioxide and nitrogen are released in the product gas at the higher value of air to fuel ratio. Decrease in the $CH_4, CO,$ and H_2 fractions in the product gas reduce the heating value of the gas. Increase in temperature from 600°C to 900°C enhances the product gas quality. The fraction of hydrogen and carbon monoxide in the product gas increased from 17% to 21% and 19% respectively. Similarly, the CGE of the product gas increased from 58% to 66.6% with increase in reactor temperature.

The CPFD model developed in this work can predict results that are good in agreement with the experimental

data. Therefore, the model can be utilized to optimize the gasification reactor.

References

C. E. Agu, L.-A. Tokheim, C. Pfeifer, and B. M. Moldestad. Behaviour of biomass particles in a bubbling fluidized bed: A comparison between wood pellets and wood chips, *Chemical Engineering Journal*, 363: 84-98, 2019. https://doi.org/10.1016/j.cej.2019.01.120

U. Arena, L. Zaccariello, and M. L. Mastellone. Fluidized bed gasification of waste-derived fuels, *Waste Management*, 30(7): 1212-1219, 2010/07/01/2010. https://doi.org/10.1016/j.wasman.2010.01.038

U. Arena and F. Di Gregorio. Gasification of a solid recovered fuel in a pilot scale fluidized bed reactor, *Fuel*, 117: 528-536, 2014. https://doi.org/10.1016/j.fuel.2013.09.044

J. Alvarez. Hydrogen production from biomass and plastic mixtures by pyrolysis-gasification, *International Journal of Hydrogen Energy*, 39(21): 10883-10891, 2014. https://doi.org/10.1016/j.ijhydene.2014.04.189

R. Beetstra, J. Nijenhuis, N. Ellis, and J. R. van Ommen. The influence of the particle size distribution on fluidized bed hydrodynamics using high-throughput experimentation, *AIChE journal*, 55(8): 2013-2023, 2009. https://doi.org/10.1002/aic.11790

J. C. Bandara, C. Jayarathna, R. Thapa, H. K. Nielsen, B. M. Moldestad, and M. S. Eikeland. Loop seals in circulating fluidized beds–Review and parametric studies using CPFD simulation, *Chemical Engineering Science*, 115917, 2020. https://doi.org/10.1016/j.ces.2020.115917

R. B. Bates et al., Steam-air blown bubbling fluidized bed biomass gasification (BFBBG): Multi-scale models and experimental validation, *AIChE Journal*, 63(5):1543-1565, 2017. https://doi.org/10.1002/aic.15666

J. Chen et al., Analysis of biomass gasification in bubbling fluidized bed with two-fluid model, *Journal of Renewable and Sustainable Energy*, 8(6): 063105, 2016. https://doi.org/10.1063/1.4967717

A. Erkiaga, G. Lopez, M. Amutio, J. Bilbao, and M. Olazar. Influence of operating conditions on the steam gasification of biomass in a conical spouted bed reactor, *Chemical engineering journal*, 237: 259-267, 2014. https://doi.org/10.1016/j.cej.2013.10.018

J. Grace and G. Sun. Influence of particle size distribution on the performance of fluidized bed reactors, *The Canadian Journal of Chemical Engineering*, 69(5): 1126-1134, 1991. https://doi.org/10.1002/cjce.5450690512

R. Jaiswal, C. E. Agu, R. K. Thapa, and B. M. E. Moldestad. Study of fluidized bed regimes using Computational Particle Fluid Dynamics, Proceedings of the 59th Conference on Simulation and Modelling, 2018. DOI:10.3384/ecp18153271

R. Jaiswal, N. C. I. Furuvik, R. K. Thapa, and B. M. E. Moldestad. Investigation of the segregation and mixing behavior of biomass in a bubbling fluidized bed reactor using a CPFD model, Proceedings of the 60th SIMS Conference on Simulation and Modelling, 2019. https://doi.org/10.3384/ecp20170164

R. Jaiswal, N. C. Furuvik, R. K. Thapa, and B. M. Moldestad. Method of identifying an operating regime in a bubbling fluidized bed gasification reactor, *International Journal of Energy Production and Management*, 5(1): 24-34, 2020. Doi: 10.2495/EQ-V5-N1-24-34

M. Niu, Y. Huang, B. Jin, and X. Wang. Simulation of Syngas Production from Municipal Solid Waste Gasification in a Bubbling Fluidized Bed Using Aspen Plus, *Industrial & Engineering Chemistry Research*, 52 (42): 14768-14775, 2013/10/23 2013. https://doi.org/10.1021/ie400026b

L. Shen, Y. Gao, and J. Xiao. Simulation of hydrogen production from biomass gasification in interconnected fluidized beds, *Biomass and bioenergy*, 32(2): 120-127, 2008. https://doi.org/10.1016/j.biombioe.2007.08.002

K.-A. Solli, R. K. Thapa, and B. M. E. Moldestad. Screening of Kinetic Rate Equations for Gasification Simulation Models, The 57th SIMS Conference on Simulation and Modelling SIMS 2016, Linköping University Electronic Press, 142: 105-112. DOI: 10.3384/ecp17142105

R. Xiao, B. Jin, H. Zhou, Z. Zhong, and M. Zhang. Air gasification of polypropylene plastic waste in fluidized bed gasifier, *Energy Conversion and Management*, 48 (3): 778-786, 2007. https://doi.org/10.1016/j.enconman.2006.09.004

J. Xie, W. Zhong, B. Jin, Y. Shao, and Y. Huang. Eulerian–Lagrangian method for three-dimensional simulation of fluidized bed coal gasification, *Advanced Powder Technology*, 24(1): 382-392, 2013. https://doi.org/10.1016/j.apt.2012.09.001

CPFD simulation of enhanced cement raw meal fluidization through mixing with coarse, inert particles

Nastaran Ahmadpour Samani [1]   Chameera K. Jayarathna[2]   Lars-Andre Tokheim[3]

[1] Department of Process, Energy and Environmental Technology, University of South-Eastern Norway, `222129@student.usn.no`

[2] SINTEF Industry, `chameera.jayarathna@sintef.no`

[3] Department of Process, Energy and Environmental Technology, University of South-Eastern Norway, `Lars.A.Tokheim@usn.no`

Abstract

In the current work, computational particle and fluid dynamics (CPFD) simulations are used to study an electrically heated bubbling fluidized bed (BFB) used as a calciner in a cement manufacturing process, applying a binary-particle fluidization system. Owing to the fine particle size (0.2 – 180 µm) of the limestone used as a raw meal in the cement kiln process, a conventional bubbling fluidized bed may be difficult to apply due to particle cohesion causing poor fluidizability of the particles smaller than 30 µm. In the current study, to enhance the fluidization of the raw meal particles, they are mixed with coarse (550 – 800 µm), inert particles. The aggregation and clustering of the fine particles will decrease due to collisions with inert coarse particles, and hence a more homogeneous distribution of raw meal particles may be achieved. The inert particles will also provide a thermal energy reservoir through their heat capacity and thereby contribute to a very stable bed temperature, which is advantageous in the control of the process. After the raw meal particles have been calcined, they have to be separated from the coarse, inert particles. This can be done by increasing the velocity of the CO_2 used for fluidization to a value sufficiently high to entrain the raw meal particles, but still sufficiently low that the coarse, inert particles are not entrained. The commercial CPFD software Barracuda was used for simulations to investigate suitable operational conditions at 1173 K, such as the particle size distribution of the inert particles and the fluidization gas velocity. The impact of gas velocity variation on the fluidization of the particle mixture was studied, and an appropriate range of velocities for the calcination and entrainment modes could be determined. The simulations revealed that mixing raw meal particles with inert coarse particles can enhance the flowability in the FB reactor indicating that it is possible to apply the concept in a full-scale calcination process.

Keywords: electrification, fluidized bed, calciner, CPFD, Barracuda, limestone

1 Introduction

Concrete is the most consumed man-made material in the world, and cement is the binder in concrete. Cement plants produce 4 billion tonnes of cement per year, corresponding to 8 percent of the global anthropogenic CO_2 emissions (Olivier et al., 2016). This makes the cement industry one of the main contributors to climate change. The main sources of CO_2 emissions in the cement industry are attributed to the chemical and thermal combustion processes in the kiln system, where clinker is produced. Clinker is the main constituent in cement.

For reducing the CO_2 emissions in the cement industry, different measures may be taken, such as 1) applying CO_2 capture methods, 2) improving existing plans to recuperate thermal and electric efficiency, 3) replacing fossil fuels with CO_2-neutral fuels, and 4) substituting clinker with alternative binders (Xu et al., 2014; Wang et al., 2018; Li et al., 2013).

Utilizing green electricity to produce clinker and applying an appropriate carbon capture technology to the reduction of CO_2 emissions of the calcination process can be a good solution to mitigate the emissions of CO_2 in the cement industry. Electrification of the calciner is most important because the calciner is the main source of CO_2 in the cement kiln system (Tokheim et al., 2019).

Fluidized bed reactors have several advantages, such as high heat and mass transfer rate, good mixing properties, and close to isothermal conditions. Because of these advantages, the fluidization concept is widely used in different engineering applications, for example reforming of hydrocarbons, gasification, aluminium production, and calcination (Wang, 2015; Dechsiri, 2004; Cocco et al., 2014; Kunii and Levenspiel, 2013).

In an electrified calciner, the heat transfer will be area-limited. Hence, it is important to provide good contact between the meal and the hot surfaces in the calciner to obtain a high overall heat transfer coefficient. A fluidized bed may be a good reactor choice as FB's are characterized by good mixing and temperature control. Hence, in the current study, an electrically

heated fluidized bed reactor is used to calcine the raw meal in the cement kiln process.

Experimental data and results from the simulation of a cold-flow lab-scale fluidized bed calciner have been compared in another paper (Ahmadpour Samani, 2020). The results revealed that owing to the vigorous inter-particular forces in the Geldart C particles (constituting about 60 % of the raw meal), it is difficult to fluidize the raw meal.

There are several methods to fluidize Geldart C particles, such as mixing them with coarser (Geldart B) particles (Kunii and Levenspiel, 2013; Tashimo et al., 1999; Kato, 1991), the mechanical agitation concept (using a chopper) (Kristensen and Schaefer, 1987; Chen et al., 2009), using Vibro-fluidization (Parikh, 2016; Chen et al., 2009), and applying a centrifugal field (Qian et al., 2001; Dave et al., 2007). In the current study, the method of mixing the small particles with inert, larger particles are investigated.

Based on experimental data and empirical observations, Geldart (Geldart, 1973) categorized particles into four types of particle behavior; A, B, C, and D, depending on particle size and density difference between solids and the fluidization gas Geldart B particles are known as a particle category having excellent fluidization properties. The large interparticle forces of Geldart C will be decreased due to collisions with inert coarse particles, and thereby a more homogeneous distribution of the fine particles will be obtained, which may facilitate fluidization. Moreover, mixing the meal with coarser particles also has the advantage of providing a thermal reservoir for the meal to be calcined.

Such a system must be operated in two different modes: 1) the calcination mode, with a relatively low fluidization gas velocity and 2) the entrainment mode, with relatively high gas velocity. When operating in the calcination mode, all particles – both raw meal and inert particles – should be maintained in the reactor. In the entrainment mode, the gas velocity is increased sufficiently that the calcined particles are entrained by the gas and pulled out of the system, whereas the inert particles remain in the FB. When designing the FB reactor, it is necessary to i) make sure all particles are calcined before they leave the FB and ii) avoid loss of inert particles from the FB.

To investigate the potential fluidization improvement without having to do experiments, computational particle and fluid dynamics (CPFD) simulations have been performed. The purpose of the CPFD investigations are:

- Find out whether it is possible to operate such a system without significant losses of inert coarse particles while at the same time having sufficient time for calcination of the meal in the reactor

- Get a clearer picture of how the different particle types behave in the FB, and the way they interact with each other

The motivation of the work is to find a reactor concept that may be used to efficiently calcine cement raw meal, with pure CO_2 as the only gaseous output from the reactor, hence avoiding the need for an expensive separation plant to capture the CO_2 from the process.

2 CPFD modelling and simulation

The commercial CPFD software Barracuda was used to study the binary-particle fluidization system, consisting of fine raw meal particles and inert coarse particles, at a temperature of 1173 K, which is the approximate calcination temperature in a full-scale system.

2.1 The CPFD method

The Barracuda Virtual Reactor, version 17.4.1, applies a numerical scheme based on the Eulerian-Lagrangian model in three dimensions for simulation of gas-particle flows. In the CPFD method, the gas phase is considered as a continuous fluid in a Eulerian grid of cells, and particles are considered as discrete Lagrangian points. The computational domain is divided into several computational parcels containing particles with the same velocity and properties. The Navier-Stokes equation is used for describing the gas phase, whereas the model of the solid phase is based on the multiphase particle-in-cell (MP-PIC) method, which is a Lagrangian description of particle motion (Andrews and O'Rourke, 1996; Snider and O'Rourke, 2011). Conservation equations for the momentum, mass, and energy are solved, including the coupling between the phases, and thereby the spatial distribution of key process parameters like temperature, pressure, and velocities in the system may be determined. Such information is necessary in order to understand the fluidization process, and that is why the CPFD scheme is widely used for simulation of gas-particle fluidized reactors (Jiang et al., 2014; Chen et al., 2013).

2.2 Simplifications

In the simulation of the binary-particle FB system in Barracuda, some simplifications have been made:

- Chemical reactions are not considered (the focus is on understanding the flow behaviour of the different particle types)

- An isothermal tubular vessel is used, i.e. no heating internals are applied in the dense bed

- The vessel is taken as adiabatic (i.e. heat loss is neglected)

The simplifications mentioned above will clearly give deviations between to the simulated results and data from an experimental system that includes heat transfer, chemical reactions and inserts. For example, chemical

reactions will change the envelope density of the particles as well as the gas flow through the bed. Furthermore, any kind of insert in the dense bed will affect the hydrodynamics of the bed and probably increase the pressure drop over the bed. process is inherently dynamic, hence dynamic (time-dependent) equations are used. However, to facilitate comparison with the results reported in another paper by the authors (Ahmadpour Samani, 2020), the same geometry is used in the current study.

The system is simulated from the initial conditions until achieving a pseudo-steady state condition.

2.3 Procedure to implement the CPFD model

In order to start developing a CPFD model, a computer-aided design (CAD[1]) model with stereolithography (STL[2]) format, which contains the geometry of the model should be prepared. For this purpose, the SolidWorks® software was used to generate the geometry of the cold-flow lab-scaled fluidized bed unit, as was also explained in another paper by the authors (Ahmadpour Samani, 2020). Next, the grid, which defines the spatial resolution for calculation of the important gas-particle flow properties including pressure, velocity, and temperature, was defined. Figure 1 represents the generated grid of the cold flow lab scaled fluidized bed unit.
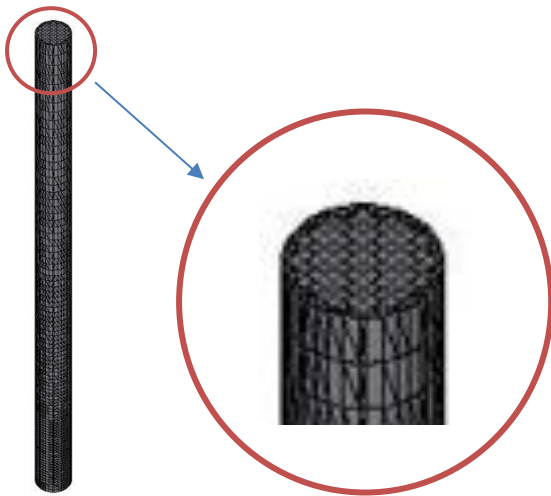

Figure 1. Grid and original CAD geometry

2.3.1 Drag model

Different built-in drag models are available in the Barracuda package, where detailed information about different drag models with references are explained in the Barracuda manual (Barracuda Virtual Reactor user manual, 2017). Since with applying the Wen & Yu drag model, previous studies of CPFD simulations of bubbling fluidized bed gasifiers have shown a good

consistency with experimental results (Jaiswal et al., 2020; Furuvik et al., 2019; Furuvik et al., 2019), the Wen-Yu drag model has been chosen for the simulations of this study.

The fluid drag force on the particle can be calculated by equation (1).

$$F_p = m_p D(u_f - u_p) \tag{1}$$

In equation (1), D defines as the drag function as shown in equation (2).

$$D = \frac{3}{8} C_d \frac{\rho_f (u_f - u_p)}{\rho_p r_p} \tag{2}$$

C_d, the drag coefficient is determined by different correlations in different drag models. The Ergun drag model is applied for dense systems, while Wen-Yu is appropriate for dilute systems (Barracuda Virtual Reactor user manual, 2017).

Equation (3) is the correlation for C_d in the Wen-Yu drag model.

$$C_d = \begin{cases} \dfrac{24\theta_f^{n_0}}{Re} & Re < 0.5 \\ (C_d)_1 & 0.5 \le Re \le 1000 \\ c_2 \theta_f^{n_0} & Re > 10000 \end{cases} \tag{3}$$

In this correlation, θ_f is the fluid volume fraction and the Re number is defined by equation (4).

$$Re = \frac{2\rho_f r_p |u_f - u_p|}{\mu_f} \tag{4}$$

$$(C_d)_1 = \frac{24}{Re} \theta_f^{n_0} (c_0 + c_1 Re^{n_1}) \tag{5}$$

The model constants are: $c_0 = 1.0, c_1 = 0.15, c_2 = 0.44, n_0 = -2.65$ and $n_1 = 0.687$

2.3.2 Initial and boundary condition as well as flux planes and transient data

Establishing the initial and boundary conditions are the succeeding step after the grid generation. In order to simulate the binary-particle fluidization system, two different types of particles are defined: $CaCO_3$ as raw meal and silica sand (SiO_2) as inert coarse particles. CO_2 was used as the fluidization gas.

The particle size of the raw meal particles ranged from 0.2 to 180 µm, whereas the inert coarse particles had a diameter in the range 550 – 800 µm.

The densities of fluidized gas and particles are listed in Table 1.

Figure 2 shows the positions of the inert particles (silica sand) and raw meal (limestone) particles in the Geldart diagram.

Figure 3 (a) shows the initial conditions in the system applied to determine a suitable fluidization velocity

---

[1] CAD software is used to create two-dimensional drawings or three-dimensional models for engineering applications.

[2] STL is a file format of a 3-D model produced by CAD software.

(more about this in section 3.2). The initial mass of $CaCO_3$ (raw meal; blue) and SiO_2 (silica sand; red) in the cylinder was 0.856 and 0.659 kg respectively. Identical layer thicknesses were used for raw meal and silica sand, but due to the difference in densities, the mass ratio was 13:10.

Figure 3 (b) shows the initial conditions in the system that corresponds to a 1:10 mass ratio between small and large particles. Initially, there were three thin layers of $CaCO_3$ (blue) and three thicker layers of silica sand (red) in the dense bed. This mass ratio was used based on the paper of (Tashimo et al., 1999). Initially, i.e. before the gas was injected, the particles were at ease in the dense bed.

Figure 3 (c) illustrates the boundaries. As for the boundary conditions, different superficial gas velocities were applied (more about this in section 3). The outlet was defined as a pressure outlet with a pressure of 1 bar, i.e., atmospheric pressure.

As the system is isothermal and adiabatic, the purpose of setting the temperature to 900 °C is to determine the right values of the thermochemical properties, such as the viscosity, which is one of the key input parameters in the drag functions.

Transient data points were used to log the system pressure at the same locations as in the cold flow lab-scaled fluidized bed unit. Figure 3(d) shows the positions of these points.

A flux plane was determined on the top of the cylinder to calculate the values of particle entrainment. Figure 3(e) shows the location of the flux plane in Barracuda.

Table 1. Temperature and density of gas and particles

Medium	Temperature [°C]	Envelope Density [kg/m³]
CO_2	900	0.45
$CaCO_3$ (raw meal particles)	20	2795
Silica sand (inert particles)	20	2200
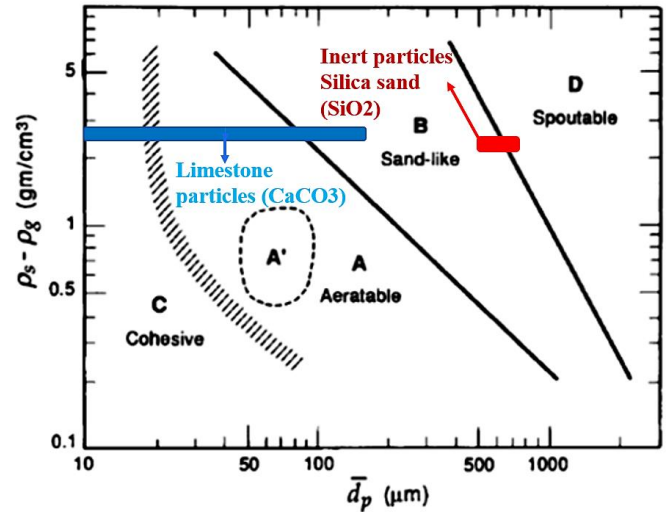
Figure 2. Geldart classification of particles for air at ambient conditions (Kunii and Levenspiel, 2013)
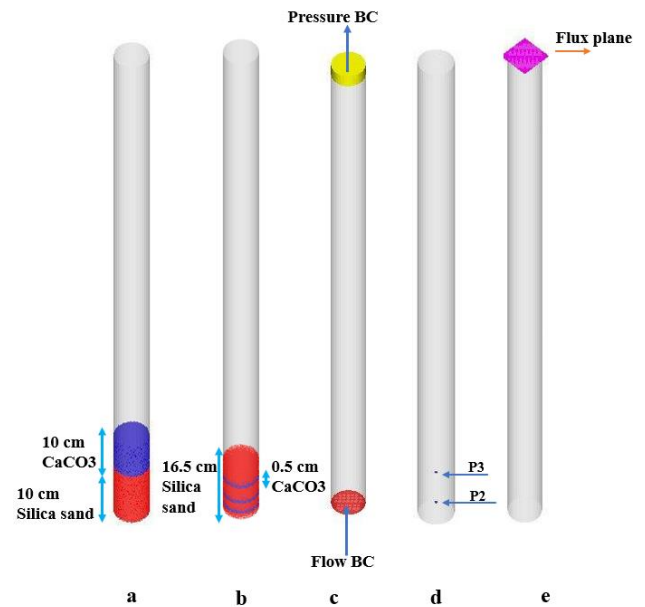
Figure 3: a, b) initial conditions, c) boundaries and d) transient points, e) Flux plane

2.3.3 Time step and duration setting

The system is simulated from t=0 (initial conditions) until a pseudo-state has been reached. Pseudo-steady state is achieved when there is no longer any net change in mass flow out of the reactor. The pseudo-steady state is checked by calculating the entrainment flux of particles through the determined flux plane (Figure 3 (e)). The simulations were conducted for 120 s of process time.

In general, for CPFD simulations, the smaller the time step, the more accurate is the solution. However, applying small time steps also results in increased calculation time. In this study, a time step of 5 ms was applied as a reasonable trade-off between accuracy and computational time.

3 Results and discussion

The minimum fluidization velocities were calculated by Wen-Yu equations for minimum fluidization velocity (Wen and Yu, 1966; Anantharaman et al., 2018). The terminal settling velocities of the particles were manually calculated by using the equations presented in (Zevenhoven and Kilpinen, 2001) and are listed in Table 2.

Table 2. The minimum fluidization velocity and terminal velocity (for the median particle size) of inert particles

D_p [μm]	V_{mf} [m/s]	V_t [m/s]
550	0.09	2.6
600	0.10	2.9
650	0.12	3.1
700	0.14	3.4
750	0.16	3.7
800	0.18	3.9

3.1 Reaction time based on the particle size

In order to calculate the reaction time of raw meal particles based on the particle size distribution, the correlations and results of the paper (Stanmore and Gilot, 2005) have been utilized. According to Stanmore et al (Stanmore and Gilot, 2005), a model based on a modified shrinking core approach (Takkinen et al., 2012) was applied for limestone particles. The shrinking core model is widely used in the description of the kinetics of limestone reactions. This model demonstrates conditions in which solid particles are being emaciated either by dissolution or reaction and, as a result, the amount of the material being consumed is "shrinking".

The calcination conversion factor can be calculated by equation (6), where k $[\frac{m^{0.6}}{s}]$ is the kinetic rate for the reaction, d_p[m] is particle diameter and t [s] is reaction time.

$$X = 1 - (1 - \frac{k}{d_p^{0.6}} t)^3 \qquad (6)$$

The value of k is given by equation (7). In this equation, T [K] is the temperature of the reaction.

$$k = 10.303 \exp\left(\frac{-10980}{T}\right) \qquad (7)$$

Hence, by applying equations (6) and (7), and assuming the values X=0.94 and T= 1173 K, the reaction time of the calcination process based on the particle sizes of the raw meal has been calculated and shown in Figure 4. It is clear from Figure 4 that the reaction time for the majority of the small raw meal particles is very short (< 1 s), hence will likely be calcined before they are entrained and leave from the top of the FB. The larger meal particles may remain in the bed during the calcination mode so that they may be sufficiently calcined before the operation is shifted to entrainment mode.
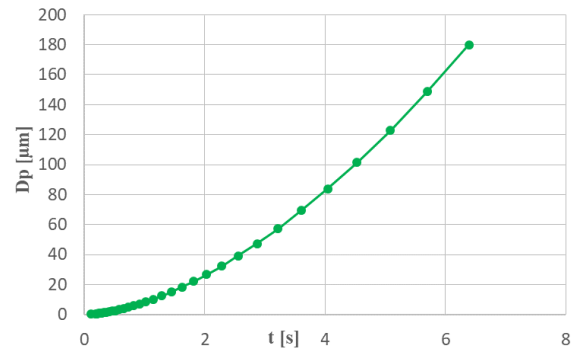
Figure 4. Reaction time for different particle sizes of limestone

3.2 The velocity range of calcination mode

Suitable operational conditions in the calcination mode must be determined. The low velocity in the calcination mode is determined in such a way that the most of the raw meal particles and all inert particles are fluidized and only a relatively small fraction of the raw meal is entrained by the gas. Barracuda is applied to find the appropriate velocity for the fluidization of the particles during the calcination mode. The velocity determined via simulation results may then be applied to design the FB reactor as an electrically heated reactor.

Figure 5 shows the minimum fluidization velocity and the terminal settling velocity of the two-particle types for different sizes (logarithmic scale). In the calcination mode, the gas velocity must be higher than the minimum fluidization velocity of the mixture but should (if possible) be lower than the terminal settling velocity of the finest particles. By using the velocity window in Figure 5, and according to the minimum fluidization velocity of the inert particles, the calcination mode velocity should be higher than the minimum fluidization velocity of the mixture, 0.13 m/s, so as to fluidize both raw meal and inert particles. Different velocities were tested in Barracuda, and the mass loss values at different velocities could be calculated using the flux plane file from the Barracuda simulation.
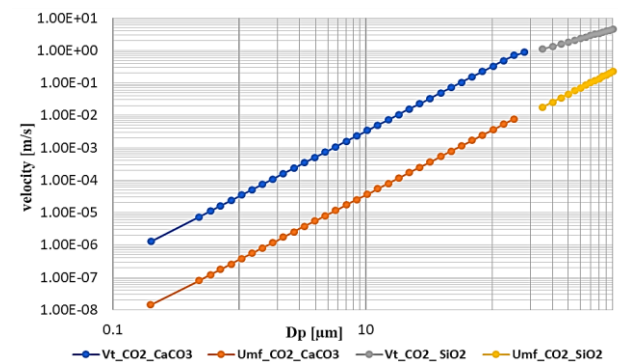
Figure 5. Velocity window for CO_2 based on different particle sizes

Table 3 shows the results for two different velocities. The simulations represented 120 s of process time. A gas velocity of 0.18 m/s, i.e. slightly higher than the minimum fluidization velocity, was selected to ensure that the particles would be fluidized. With this velocity, only 7.6 % of the raw meal was entrained, and this velocity is to be used for the design of the FB reactor.

Table 3. The values of entrained mass in the calcination mode for a mass ratio of 1:1

| Case	Velocity [m/s]	CaCO$_3$ loss [kg]	CaCO$_3$ loss [%]	SiO$_2$ loss [kg]	SiO$_2$ loss [%]
1	0.18	0.065	7.6	0	0
2	0.22	0.179	21	0	0

Based on the results discussed above, the simulations with the ratio 1:10 were conducted at a velocity of 0.18 m/s. The initial mass of raw meal was 0.128 kg, and at this velocity, only 2.8 % of the meal left the cylinder. The results were calculated by using the data of the flux plane file in Barracuda and are listed in Table 4.

Table 4. The value of mass entrained at 0.18 m/s with a mass ratio of 1:10

Velocity [m/s]	CaCO$_3$ loss [kg]	CaCO$_3$ loss [%]	SiO$_2$ loss [kg]	SiO$_2$ loss [%]
0.18	0.004 kg	2.8	0	0

Figure 6 depicts the fluidization and mixing of the raw meal (limestone, blue) and inert (silica sand, red) particles in the FB reactor. In the case with the ratio 1:10, the contact between raw meal and inert particles and heat transfer is increased. The inert particles also provide a thermal energy reservoir through their heat capacity and thereby contribute to a very stable bed temperature, which is advantageous in the control of the process. Hence, in the design of the FB reactor, a mass ratio of 1:10 may be suitable (although other mass ratios, somewhat higher or lower may also be considered).

Figure 6 also indicates that it takes as much as 30 seconds before most of the raw meal particles have reached the top of the dense bed. Considering the required time for calcination shown in Figure 4 (< 7 seconds), this indicates that sufficient heat will have been transferred from the large, inert sand particles to the small raw meal particles so that it is likely that they will have been sufficiently calcined during the contact period.
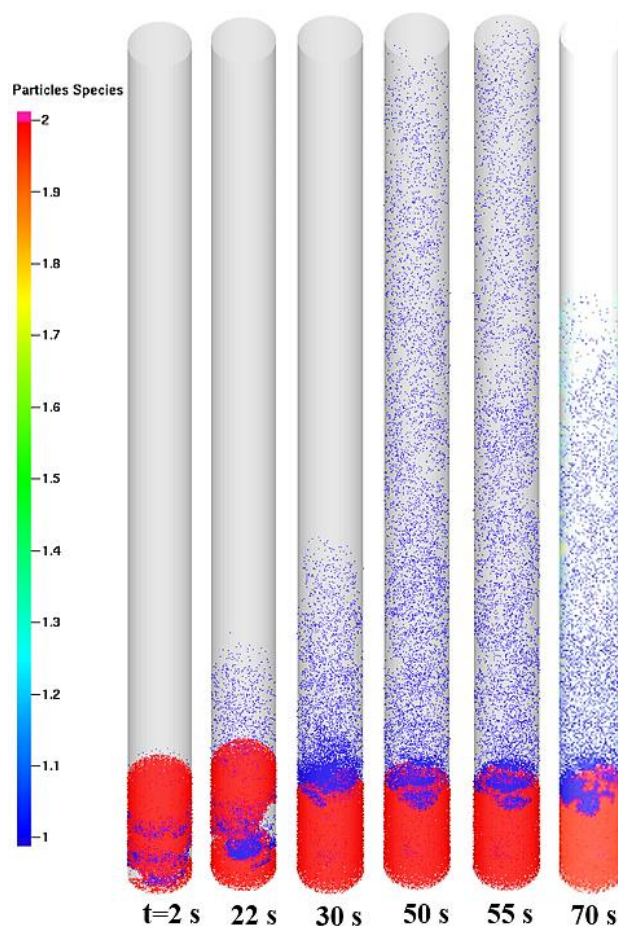
Figure 6. Distribution of limestone and silica sand particles with the ratio 1/10 at different periods, v=0.18m/s

3.3 The appropriate gas velocity for entrainment stage

When the calcination process is close to complete and the calcination degree reaches 94%, the gas velocity will be increased to a high value, so that the calcined meal particles (but not the inert coarse particles) will be entrained from the bed. The velocity of the entrainment stage must be higher than the terminal settling velocity of the fine raw meal particles, but lower than the terminal settling velocity of the inert coarse particles. According to the data in Figure 5, the velocity can be between 0.9 and 3.2 m/s.

In order to find an appropriate gas velocity in the entrainment stage, different cases with the same initial height of raw meal and inert particles have been simulated. In Table 5, two cases with velocities of 1.5 m/s and 3 m/s are compared. With the velocity of 3 m/s, almost all raw meal particles were entrained from the bed, but no inert particles left the bed. Therefore, a velocity of 3 m/s in the entrainment stage is seen as appropriate. The data has been obtained by applying the flux plane file of Barracuda. The simulations were conducted for 120 s.

Table 5. The values of the entrained mass of raw meal and inert particles in the entrainment stage

Case	Velocity [m/s]	$CaCO_3$ loss [kg]	$CaCO_3$ loss [%]	SiO_2 loss [kg]	SiO_2 loss [%]
1	1.5	0.659	77	0	0
2	3.0	0.850	99	0	0
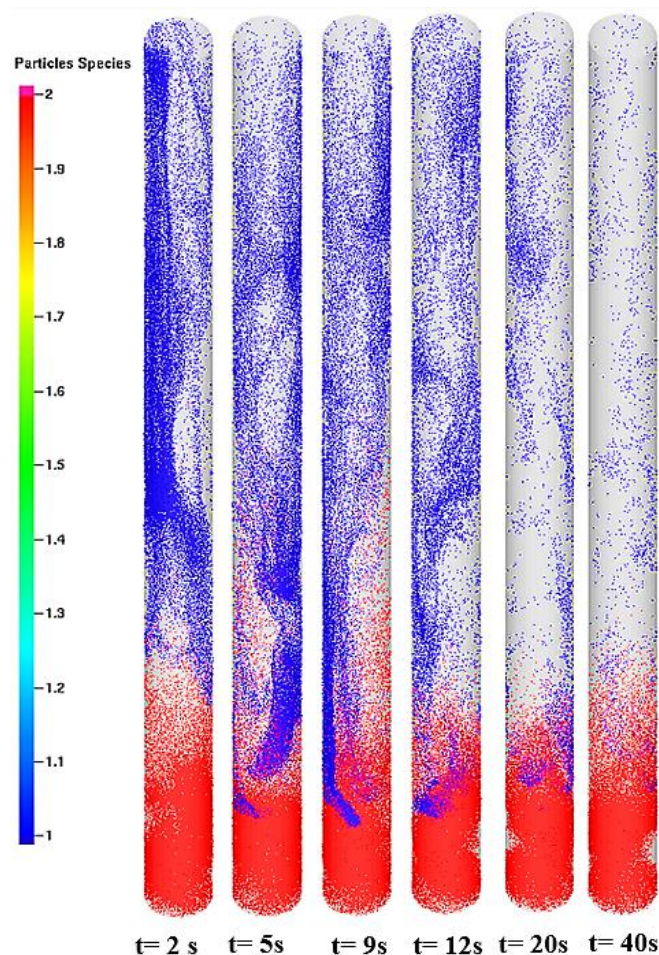
Figure 7. The distribution of raw meal and inert particles in the entrainment stage

As can be seen in Figure 7, at the gas velocity of 3 m/s, the raw meal particles left the bed, but the inert particles remained in the bed. After entrainment of the calcined meal, the bed is ready for a new calcination period.

4   Conclusion

To enhance the fluidization of the raw meal particles (0.2-180 µm; i.e. Geldart C), mixing them with inert coarse particles (550-800 µm; i.e. Geldart B) was proposed.

Such mixing tends to counteract the interparticular forces in the Geldart C powder. Another advantage is that the Geldart B particles provide a thermal energy reservoir through their heat capacity and thereby contribute to a stable bed temperature.

Barracuda was used for CPFD simulations to determine a suitable operational gas velocity at 1173 K. The inert coarse particles were found to have an average minimum fluidization velocity of 0.129 m/s and a terminal settling velocity of 3.24 m/s. The appropriate velocities for the calcination and entrainment modes were determined as 0.18 m/s and 3 m/s for a mass ratio of 1:10 between raw meal and sand particles. Based on literature results, this is believed to be a suitable mass ratio.

The simulations illustrated that at the velocity of 0.18 m/s (calcination velocity), with the 1:10 mass ratio, only 2.8 % of the raw meal particles were lost from the bed during the calcination period. At the velocity of 3 m/s (entrainment velocity), 99 % of the raw meal particles were entrained, while no inert particles were entrained. Hence, it is possible to operate the system with an appropriate separation efficiency during the entrainment period.

The calculated reaction time of the limestone particles revealed that limestone particles in the diameter range of 0.2 – 70 µm were calcined in the time range 0.1 – 3.7 s. Therefore, they are likely to be calcined before leaving the reactor in the calcination mode.

The simulations indicate that it is feasible to enhance the flowability of raw meal in the FB reactor by mixing them with the inert silica sand particles. The results also indicate that it is possible to obtain sufficient calcination of the raw meal particles during the calcination period and a high degree of separation during the entrainment period. This means that the concept should be possible to apply in a full-scale calcination process.

In future work, the calcination process will be simulated for a full-scale system including heat transfer and chemical reactions.

References

N. Ahmadpour Samani, Ch. K. Jayarathna, and L.A. Tokheim. Fluidized bed calcination of cement raw meal: Laboratory experiments and CPFD simulations, In *Proceedings*, *61th SIMS conference, 22-24 September,Finland,* 2020.

A. Anantharaman, R. A. Cocco, and J. W. Chew. Evaluation of correlations for minimum fluidization velocity (Umf) in gas-solid fluidization, *Powder technology*, 323: 454-85, 2018.

M. J. Andrews and P. J. O'Rourke. The multiphase particle-in-cell (MP-PIC) method for dense particulate flows, *International Journal of Multiphase Flow*, 22: 379-402, 1996.

C. Chen, J. Werther, S. Heinrich, H. Y. Qi, and E. Ulrich. CPFD simulation of circulating fluidized bed risers. *Powder technology Hartge*, 235: 238-47. 2013.

Y. Chen, J. Yang, R. N. Dave, and R. Pfeffer. Granulation of cohesive Geldart group C powders in a Mini-Glatt fluidized bed by pre-coating with nanoparticles, *Powder technology*, 191: 206-17, 2009.

R. Cocco, S. R. Karri, and T. Knowlton. Introduction to fluidization, *Chem. Eng. Prog*, 110: 21-29, 2014.

Barracuda Virtual Reactor user manual. Houston, USA, 2017. https://cpfd-software.com

R. Dave, R. Pfeffer, J. Yang, and Y. Chen. Dry coating and downstream processing of cohesive powders. *In.: Google Patents*, 2007.

Ch. Dechsiri. Particle transport in fluidized beds, *University Medical Center Groningen, University of Groningen,* 2004.

N. C. I. Furuvik, R. Jaiswal, and B. M. E. Moldestad. Flow behavior in an agglomerated fluidized bed gasifier, *The International Journal of Energy and Environment*, 10 (2), 55-64, 2019.

N. C. I. Furuvik, R. Jaiswal, R. K. Thapa, and B. M. E. Moldestad. CPFD model for prediction of flow behavior in an agglomerated fluidized bed gasifier, *Int. J. of Energy Prod. & Mgmt*, 2019.

D. Geldart. Types of gas fluidization, *Powder technology*, 7: 285-92, 1973.

R. Jaiswal, N.C. I. Furuvik, R. K. Thapa, and B. M. E. Moldestad. Method of identifying an operating regime in a bubbling fluidized bed gasification reactor, *International Journal of Energy Production and Management*, 5: 24-34, 2020.

Y. Jiang, G. Qiu, and J. Haigang. Modelling and experimental investigation of the full-loop gas–solid flow in a circulating fluidized bed with six cyclone separators, *Chemical Engineering Science Wang*. 109: 85-97, 2014

K. Kato. Residence time distribution of fine particles in a powder-particle fluidized bed, *Kagaku Kogaku Ronbunshu*, 17: 970-75, 1991.

H. G. Kristensen and T. Schaefer. Granulation: a review on pharmaceutical wet-granulation, *Drug development and industrial pharmacy*, 13: 803-72, 1987.

D. Kunii and O. Levenspiel. *Fluidization engineering*. Elsevier, 2013.

J. Li, P. Tharakan, D. Macdonald, and J. Xi. Technological, economic and financial prospects of carbon dioxide capture in the cement industry, *Energy Policy Liang*, 61: 1377-87, 2013.

J. Olivier, G. J. Maenhout, M. Muntean, and J. Peters. Trends in global CO2 emissions; 2014 Report, PBL Netherlands Environmental Assessment Agency; *European Commission, Joint Research Centre, Ispra, Italy, 2016*.

D. M. Parikh. *Handbook of pharmaceutical granulation technology*, CRC Press. 2016.

G. H. Qian, I. Bágyi, I. W. Burdick, R. Pfeffer, H. Shaw, and J. G. Stevens. Gas–solid fluidization in a centrifugal field, *AIChE Journal*, 47: 1022-34. 2001.

D. M. Snider and P. J. O'Rourke. The multiphase particle-in-cell (MP-PIC) method for dense particle flow. *Computational Gas-Solids Flows and Reacting Systems: Theory, Methods and Practice* (IGI Global), 2011.

B. Stanmore and P. Gilot. calcination and carbonation of limestone during thermal cycling for CO2 sequestration, *Fuel processing technology*, 86: 1707-43, 2005.

S. Takkinen, J. Saastamoinen, and T. Hyppänen. Heat and mass transfer in calcination of limestone particles, *AIChE Journal*, 58: 2563-72, 2012.

T. Tashimo, T. Suto, J. Murota, and K. Kato. Calcination of fine limestone particles by a powder-particle fluidized bed, *Journal of chemical engineering of Japan*, 32: 374-78, 1999.

L. A. Tokheim, A. Mathisen, L. E. Øi, Ch. Jayarathna, N. H. Eldrup, and T. Gautestad. Combined calcination and CO2 capture in cement clinker production by use of electrical energy. *TCCS–10. CO2 Capture, Transport and Storage. Trondheim 17th–19th June 2019. Selected papers from te 10th International Trondheim CCS Conference*. pages 101-109, 2019.

H. Wang, W. Chen, and J. Jingcheng. Low carbon transition of global building sector under 2-and 1.5-degree targets, *Applied energy*. 2018222: 148-57, 2018.

M. Wang. *Industrial tomography: systems and applications*. Elsevier. 2015

C. Wen and Y. Yu. A generalized method for predicting the minimum fluidization velocity, *AIChE Journal*, 12: 610-12, 1966.

J. H. Xu, T. Fleiter, Y. Fan, and J. Wolfgang. CO2 emissions reduction potential in China's cement industry compared to IEA's Cement Technology Roadmap up to 2050, *Applied Energy Eichhammer*, 130: 592-602, 2014.

R. Zevenhoven and P. Kilpinen. *Control of pollutants in flue gases and fuel gases*. Helsinki University of Technology Espoo, Finland, 2001.

Fluidized bed calcination of cement raw meal: Laboratory experiments and CPFD simulations

Nastaran Ahmadpour Samani [1]   Chameera K. Jayarathna[2]   Lars-Andre Tokheim[3]

[1] Department of Process, Energy and Environmental Technology, University of South-Eastern Norway, `222129@student.usn.no`

[2] SINTEF Industry, `chameera.jayarathna@sintef.no`

[3] Department of Process, Energy and Environmental Technology, University of South-Eastern Norway, `Lars.A.Tokheim@usn.no`

Abstract

The chemical and thermal processes associated with the decarbonation and fuel combustion in the cement kiln process produce a large amount of carbon dioxide (CO_2) contributing with around 8 % of the global CO_2 emissions. Utilizing green electricity instead of fossil fuels to decarbonate the raw meal in the calciner can eliminate the CO_2 emissions produced through fuel combustion and also provide a basis for simple capture of the CO_2 generated through calcination because CO_2 is the only gaseous product exiting from the electrified calciner. In the current work, an electrically heated fluidized bed (FB) reactor is being developed to calcine the raw meal. The FB may replace the traditional entrainment calciner used in many plants. The purpose is to enable efficient indirect heat transfer in the bubbling bed and hence obtain pure CO_2 as the gaseous product from the calciner. The minimum fluidization velocity and pressure drop of the particle bed are important characteristics in the design of a bubbling fluidized bed, and these have been measured in a cold-flow lab-scale fluidized bed unit with a bed height of 0.21 m and a circular cross-sectional area of 55 cm². The particle size distribution of the meal ranged from 0.2 – 180 µm, with a median particle size of 21 µm. The experimental results revealed that the regular cement raw meal is difficult to fluidize due to the large fraction of Geldart C particles in the meal (approximately 60%). Based on experimental observation, this may be explained by inter-particular electrostatic forces forming particle clusters. The fluidization process has also been simulated with the commercial computational particle and fluid dynamics (CPFD) software Barracuda® (version 17.4.1). The purpose of using CPFD was to be able to simulate the process at cold-flow conditions and then, based on this, simulate the process at large-scale hot-flow conditions. The simulation results complied quite well with the lab-scale experiments and confirmed the difficult fluidization of the meal.

Keywords: bubbling fluidized bed, calciner, electrification, CPFD, Barracuda, Geldart C, limestone

1 Introduction

Cement is a key constituent in concrete, the most widely used building material in the world. The cement industry is one of the main contributors to climate change by producing 8 % of the global CO_2 emissions. Hence, strict carbon capture mitigation strategies are needed in the cement industry to comply with the Paris agreement on climate change.

Clinker is the main constituent in cement, and in the clinker production process, there are two major contributions to emissions of CO_2: 1) The raw meal contains typically 75-80 % calcium carbonate ($CaCO_3$), and during the calcination of the raw meal, $CaCO_3$ will decompose into lime (CaO) and CO_2: $CaCO_3 \rightarrow CaO + CO_2$. In a modern cement kiln, the calcination process typically contributes to 65% of the CO_2 emissions. 2) As the calcination is an endothermic process, fuel combustion is used to provide thermal energy, contributing to about 35% of CO_2 emissions (Andrew, 2018; Tokheim et al., 2019).

Utilizing green electricity instead of fossil fuels to decarbonate the raw meal in the cement kiln process can eliminate the CO_2 emissions produced through fuel combustion and also provide a basis for simple capture of the CO_2 generated through calcination, as CO_2 is the only gaseous product exiting from the electrified calciner (Tokheim et al., 2019).

Different reactors can be used as an electrified calciner. A bubbling fluidized bed (BFB) provides good mixing, hence giving efficient heat transfer and good temperature control. The BFB may operate with almost isothermal conditions, and the thermal reservoir provided by the bed prevents abrupt process changes. Hence, an electrically heated bubbling fluidized bed reactor is used to calcine the raw meal in the current study.

The BFB will replace the traditional entrainment calciner used in many cement plants. The purpose is to

enable efficient indirect heat transfer in the bubbling bed and hence obtain pure CO_2 as the gaseous product from the calciner.

2 Methods

In order to design an electrically heated BFB reactor to calcine the raw meal, two approaches have been combined: 1) Lab-scale experiments have been performed in order to determine some key design variables for the BFB. 2) CPFD simulations have been used to simulate the process in the lab-scale unit. The experimental part is further described in section 4, whereas the simulation part is described in section 5.

3 Theoretical considerations

Raw meal from Norcem Brevik (a Norwegian cement plant) was used as the basis for the experiments, simulations, and design work. The meal had a particle size distribution ranging from 0.2 to 180 µm and a median particle size of 21 µm. 59 % of the particles were below 30 µm, i.e. being Geldart C particles (Kunii and Levenspiel, 2013), whereas 30 % and 11 % were Geldart A and B particles, respectively.

Geldart C particles are difficult to fluidize and tend to grow as a plug of particles. When exposed to the fluidization gas, cracks, channels, or so-called rat holes are formed, and the particles are not properly fluidized, especially if the diameter of the bed is large. The fluidization behavior of Geldart C powders is due to the vigorous inter-particle forces. When the surface-to-volume ratio of the particles increases, the interparticle forces get larger, and consequently the distance between the particles is reduced. The cohesive forces become greater than the particle gravity and the hydrodynamic forces applied by fluidization gas around the particle (Kunii and Levenspiel, 2013; Chen et al., 2009).

With a larger particle size, it could have been possible to fluidize the meal well. However, use a coarser meal is not an option as small particles are required to obtain the right clinker quality.

One of the approaches that can be applied to facilitate fluidization of Geldart C particles is mixing them with Geldart B particles. According to Kunii and Levenspiel (Kunii and Levenspiel, 2013), "One way of processing these solids is to introduce them into a bed of the same material but of larger size, preferably Geldart B. Even though the fines are very small, they are not entrained immediately but may stay in the bed an average of several minutes. This usually is long enough for a physical or chemical transformation of these solids."

Since the meal does contain some Geldart B (and A) particles, the experimental tests have been applied to test the flowability of the meal.

4 Experimental work

To investigate the fluidizability of the raw meal of limestone particles, some tests have been conducted by a cold-flow lab-scale fluidized bed unit.

4.1 Experimental procedure

Figure 1 shows a cold-flow lab-scale fluidized bed unit with a height of 1.5 m and a circular cross-sectional area of 55 cm². The gas flow rate is controlled by the mass flow rate controller, and the pressure is measured by pressure transmitters mounted at different axial positions in the cylinder wall (P1 to P9). The distance between two adjacent transmitters is 10 cm. The pressure and volume flow rates are recorded through a LabVIEW program.

Table 1 shows some key experimental data.
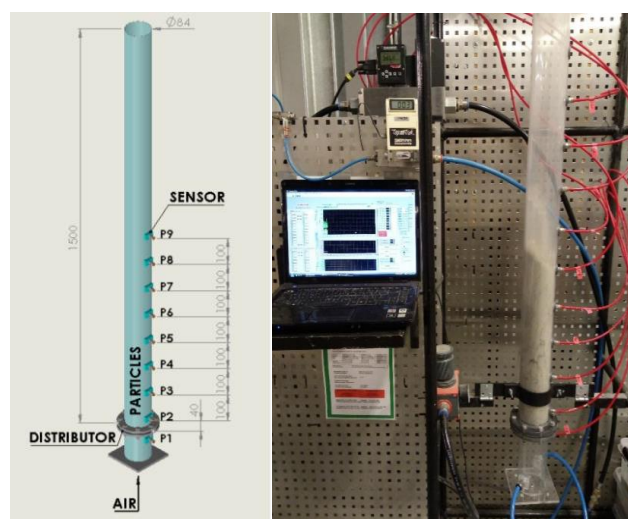
Figure 1. Cold flow lab-scaled fluidized bed unit.

Table 1: Experimental data (fluidization air properties taken from (Incropera et al., 2006)).

Total bed height	1.5 m
Bed diameter	84 mm
Particle size	0.2-180 µm
Median particle size	21 µm
Particle density	2795 kg/m³
Initial dense bed height	0.21 m
Air density	1.1707 kg/m³
Air viscosity	1.836·10⁻⁵ Pa·s

4.2 Experimental results

As described above, close to 60 percent of particles belong to the cohesive Geldart C powder, and Figure 2 indeed confirms that proper fluidization does not occur. When the gas flow rate is increased, rat holes and cracks

are formed, and the particles are not fluidized. Instead, the particles have a tendency to form clusters of particles probably due to electrostatic forces, and the particle-particle interactions prevent proper fluidization. Ratholes and channeling effects are observed in the figures. The higher the superficial gas velocity, the higher number of rat holes. With increasing gas velocity, the height of the bed is almost doubled. At high gas flow rates, a significant fraction of the particles is entrained and leave the cylinder, whereas larger particles at the bottom of the bed form clusters. Consequently, in order to fluidize this particle size distribution of raw meal, other methods should be applied.
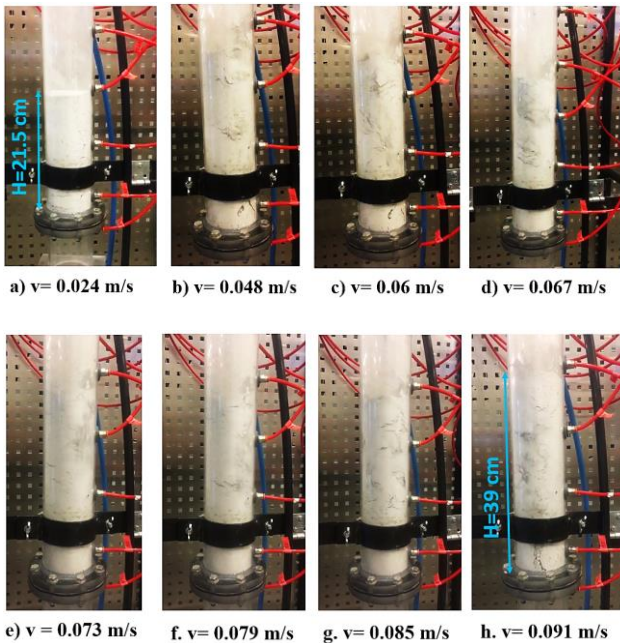
Figure 2. Snapshots from experimental tests with different superficial gas velocities

A diagram of pressure drop over the bed versus superficial gas velocity is shown in Figure 3. Since the fluidization did not happen in the bed, the minimum fluidization velocity could not be determined experimentally by this diagram. According to the observations, rat holes are starting to form when the superficial gas velocity reaches 0.036 m/s, which explains the pressure drop decline.
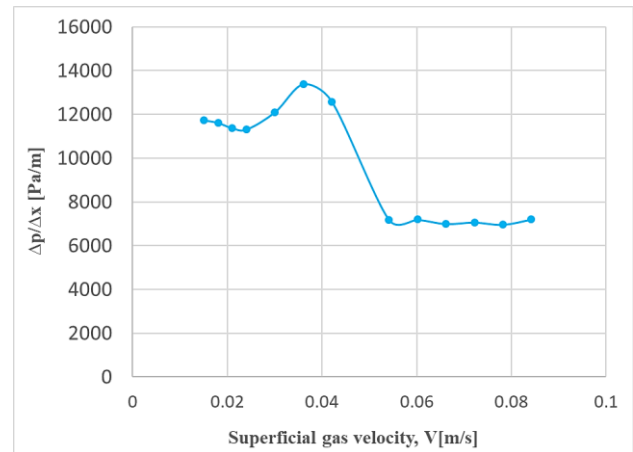

Figure 3. Experimental results for pressure drop vs superficial gas velocity.

5 Simulations

Computational Particle Fluid Dynamics (CPFD) modelling is applied to simulate the experimental system, aiming at finding consistency between the experimental results and the simulation results.

5.1 The CPFD method

CPFD is a numerical method based on the Eulerian-Lagrangian method to simulate a large-scale multiphase (particle-fluid) flow system in three dimensions by adopting the multiphase particle-in-cell (MP-PIC) method and the particle parceling algorithm (Andrews and O'Rourke, 1996; Snider and O'Rourke, 2011). The Eulerian-Lagrangian method uses a continuum model for the fluid phase and a Lagrangian method is applied for the particle phase. This gives an appropriate numerical solution for a wide range of particle sizes, shapes, and velocities. The Navier-Stokes equation with coupling between the discrete particles is applied for the fluid phase. The direct element method (DEM) fits into the Lagrangian method to solve the particle phase (Chen et al., 2013; Jiang et al., 2014).

5.2 Simulation set-up

In this project, to find the minimum fluidization velocity as the main characteristic of the design of bubbling fluidized bed reactor, the flow of particles in a bubbling fluidized bed has been simulated with different flow rates. The goal is to be able to simulate the physical process at cold-flow conditions and, based on this, be able to use the CPFD model to simulate an upscaled high-temperature calcination process and hence predict the behavior of the full-scale process.

A computer-aided design (CAD) model of the BFB geometry with a stereolithography (STL) format has been prepared by applying SolidWorks.

Grid generation is the base for all simulations since it determines the spatial resolution for calculating the flow properties of the particle-gas flow. The resolution

should be sufficiently high, so that the particle-fluid dynamics can be calculated with sufficient accuracy. However, increasing the number of cells leads to an increase in the time of the calculation. Hence, a suitable trade-off between accuracy and simulation time should be found. Besides, to achieve a stable and efficient simulation, having close to uniform cell sizes in the grid is important. Figure 4 illustrates the generated grid on the model.

Figure 4. (a) Grid and original CAD geometry and (b) Original model geometry

The drag force (F_p) acting on a particle through a fluid flow can be calculated by equation (1). In this equation, m_p is the particle mass, u_f is the fluid velocity, u_p is the particle velocity and D is a drag function, shown in equation (2).

$$F_p = m_p D (u_f - u_p) \quad (1)$$

$$D = \frac{3}{8} C_d \frac{\rho_f (u_f - u_p)}{\rho_p r_p} \quad (2)$$

There are several drag models in Barracuda, and each of them provides specific correlations for the drag coefficient, C_d. The Wen-Yu drag model is appropriate for the dilute systems, and equation (3) shows the correlation for C_d as a function of Reynolds number (equation (4)). θ_f is the fluid volume fraction.

$$C_d = \begin{cases} \dfrac{24 \theta_f^{n_0}}{Re} & Re < 0.5 \\ (C_d)_1 & 0.5 \leq Re \leq 1000 \\ c_2 \theta_f^{n_0} & Re > 10000 \end{cases} \quad (3)$$

$$Re = \frac{2 \rho_f r_p |u_f - u_p|}{\mu_f} \quad (4)$$

$$(C_d)_1 = \frac{24}{Re} \theta_f^{n_0} (c_0 + c_1 Re^{n_1}) \quad (5)$$

The model constants are:

$$c_0 = 1.0, c_1 = 0.15, c_2 = 0.44, n_0 = -2.65 \ and$$

$$n_1 = 0.687$$

The Ergun model, shown in equation (6), is suitable for higher packing fractions. The Wen-Yu/Ergun blend drag model is a combination of the Wen-Yu and Ergun drag models and can be applied for dilute as well as dense systems.

$$D = 0.5 (\frac{c_1 \theta_p}{\theta_f Re} + c_0) \frac{\rho_f |u_f - u_p|}{r_p \rho_p} \quad (6)$$

Here, c_0=2 and c_1=180 (Barracuda Virtual Reactor user manual, 2017).

Approximately 60% of the raw meal particles belong to Geldart C, and there is not any specific drag model recommended for this group of particles (Jayarathna et al., 2014; Jayarathna et al., 2017). However, the Wen-Yu drag model was used to simulate different cases in the current study.

Figure 5 shows the initial condition (a) and the boundary conditions (b). The initial height of limestone particles was set to 21 cm. In Figure 5 (b), the inlet flow position, and the outflow position have been specified with red and yellow colours, respectively.
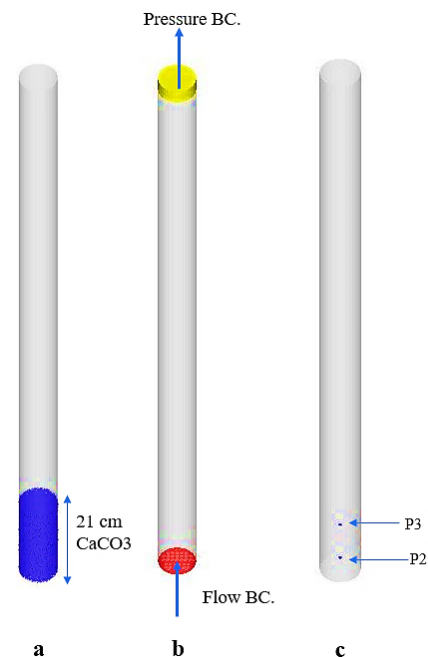
Figure 5. Initial condition (a), boundary condition (b) and transient points (c)

The time step should be chosen so as to achieve pseudo-steady-state conditions as well as give reasonable total calculation time. To make sure that (pseudo) steady-state conditions would be reached, based on experience, and also including a safety factor, the simulations were conducted for 60 s of process time for all cases.

For data exploration and analysis, flux planes may be defined and transient data from the simulations can be saved in text files. The date in the text files provide a basis for comparison between simulations and experimental data. As can be seen in Figure 5 (c), transient points were determined in the positions of P2 and P3 pressure transmitters. By using the text files and collecting the value of pressures in the two positions, the minimum fluidization velocity can be calculated.

5.3 Simulation results and comparison with experimental results

The simulation results for ambient temperature and a velocity equal to 0.08 m/s are presented in Figure 6. A diagram of pressure drop versus superficial gas velocity was also plotted, see Figure 7. The figure indicates that the simulation results comply well with the experimental data. After increasing the air velocity, it does penetrate into the bed, and the pressure drop increases. However, with rathole formation, the pressure drop starts to decrease. The figures illustrate that a pseudo-steady-state condition is obtained after around 20 seconds, hence demonstrating that the selected superficial gas velocity is likely adequate.

Both experimental results and simulations reveal that the fine Geldart C particles cannot be fluidized by conventional methods of fluidization. Other methods may be applied in order to fluidize this group of particles, such as mixing with coarse (Geldart B) particles, using flow conditioners, mechanical vibration, sound-assisted fluidization, fluidization with magnetic/electric fields, pulse fluidization, or centrifugal fluidization (Kristensen and Schaefer, 1987; Chen et al., 2009; Parikh, 2016)
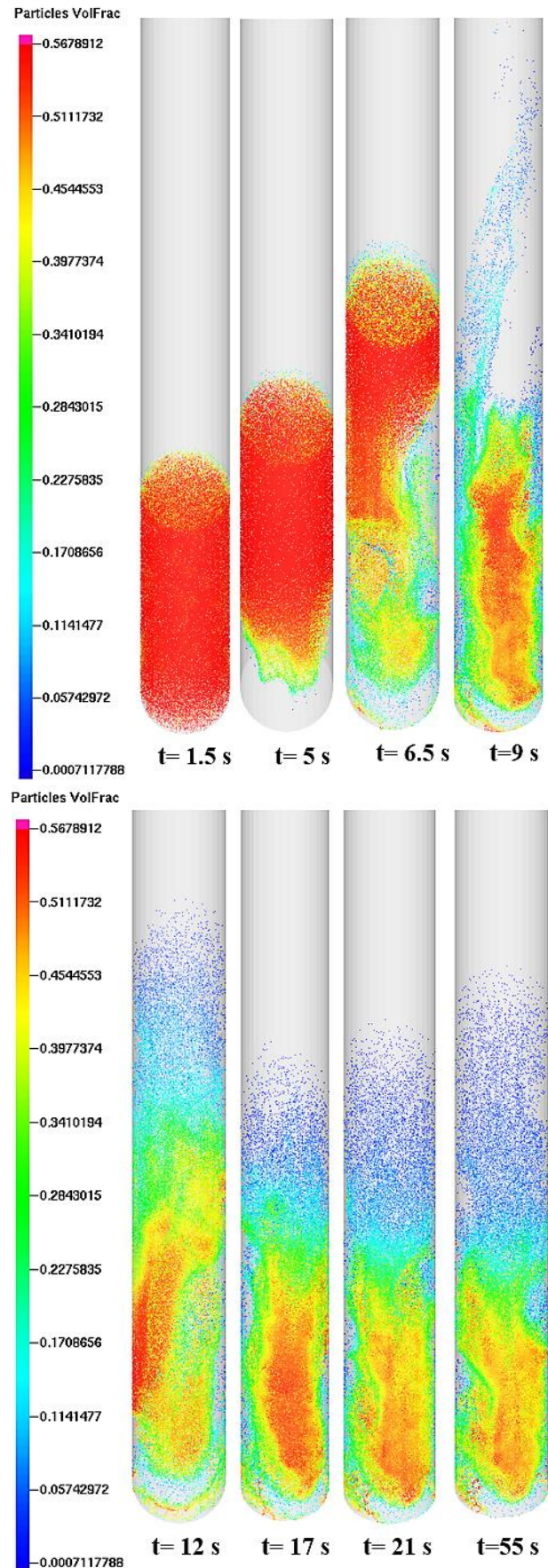
Figure 6. Simulation results of particle volume fraction of limestone
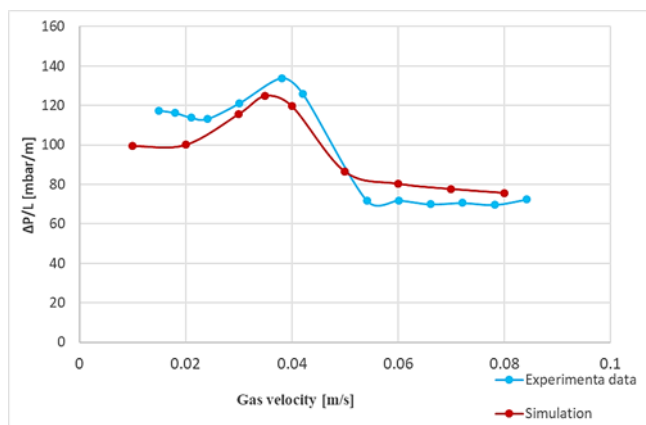
Figure 7. The pressure drop versus velocity for experimental and simulation results.

6 Discussion

According to (Geldart, 1973; Kunii and Levenspiel, 2013; Ram, 2013), due to the large inter-particular forces between Geldart C particles, they are difficult to fluidize. The results of experiments and simulations confirmed this. Therefore, although the meal of limestone consists of Geldart A and B, the fine powders of Geldart C have dominated the meal (approximately 60% of the meal) and there are not enough large particles to give a significant improvement in fluidization.

The minimum fluidization velocities of larger particles (120-500 µm) have previously been calculated by Barracuda simulations and determined in experiments (Jayarathna et al., 2014) and showed reasonably good agreement with each other and with other literature data (Yang et al., 2004). The deviation between simulated and measured data varies between 1% and 15 %, and the root means the square error is 10%. These studies indicated that fluidization is much easier for Geldart A and B particles.

One of the approaches that can be applied to improve the flowability of Geldart C particles is mixing them with a large fraction of Geldart B particles, as explained in the literature (Kunii and Levenspiel, 2013) and as shown in experiments with small limestone particles (Tashimo et al., 1999; Kato, 1991).

7 Conclusion

Although there were some fractions of larger particles of Geldart A and B in the raw meal, the results of the experiments of the cold-flow lab-scale BFB unit revealed that regular cement raw meal is difficult to fluidize. This is likely due to the large fraction of Geldart C particles in the meal (approximately 60%). Owing to the fine particle sizes of the raw meal, there are strong cohesive forces between the particles, and this prevents fludization.

The results confirm the results published in the literature about Geldart C particles. Hence, a conventional bubbling fluidized bed is difficult to fluidize Geldart C particles. The pressure drop profile and the minimum fluidization velocity determined by CPFD simulation compared quite well with the experimental results (RMSE 10 %).

In future work, the mixing of the cement raw meal and a rather high fraction of inert Geldart B particles will be investigated to find out whether this could be a way to solve the fluidization problem of the regular cement raw meal.

References

R. M. Andrew. Global CO2 emissions from cement production. *Earth Syst. Sci. Data.* 10: 195-217, 2018.

M. J. Andrews and P. J. O'Rourke. The multiphase particle-in-cell (MP-PIC) method for dense particulate flows. *International Journal of Multiphase Flow*, 22: 379-402. 1996.

C. Chen, W. Joachim, H. Stefan, Q. Hai Y, and H. Ernst-Ulrich. CPFD simulation of circulating fluidized bed risers, *Powder technology*, 235: 238-47, 2013.

Y. Chen, J. Yang, R. N. Dave, and R. Pfeffer. Granulation of cohesive Geldart group C powders in a Mini-Glatt fluidized bed by pre-coating with nanoparticles, *Powder technology*, 191: 206-17. 2009.

Barracuda Virtual Reactor user manual. Houston, USA, 2017. *https://cpfd-software.com*

D. Geldart. Types of gas fluidization, *Powder technology*, 7: 285-92, 1973.

F. P. Incropera, D. P. DeWitt, T. L. Bergman, and A. S. Lavine. *Fundamentals of Heat and Mass Transfer*, Sixth edn John Wiley and Sons, New Jersey, USA, 2006.

C. K. Jayarathna, B. M. Halvorsen, and L. A. Tokheim. Experimental and theoretical study of minimum fluidization velocity and void fraction of a limestone based CO2 sorbent, *Energy Procedia*, 63: 1432-45. 2014.

C. K. Jayarathna, B. M. Moldestad, and L. A. Tokheim. Validation of results from Barracuda® CFD modelling to predict the minimum fluidization velocity and the pressure drop of Geldart A particles, *Proceedings of the 58th SIMS September 25th - 27th, Reykjavik, Iceland*, 2017. doi: 10.3384/ecp1713876.

Y. Jiang, G. Qiu, and H. Wang. Modelling and experimental investigation of the full-loop gas–solid flow in a circulating fluidized bed with six cyclone separators, *Chemical engineering science*, 109: 85-97, 2014.

K. Kato. Residence time distribution of fine particles in a powder-particle fluidized bed, *Kagaku Kogaku Ronbunshu*, 17: 970-75, 1991.

H. Kristensen and T. Schaefer. Granulation: a review on pharmaceutical wet-granulation, *Drug development and industrial pharmacy*, 13: 803-72, 1987.

D. Kunii and O. Levenspiel. *Fluidization engineering*, Elsevier, 2013.

D. M. Parikh, *Handbook of pharmaceutical granulation technology*, CRC Press, 2016.

D. K. Ram. The determination of minimum bubbling velocity, minimum fluidization velocity and fluidization index of fine powders (hematite) using gas-solid tapered beds,

International Journal of Science and Research, 2: 287-93, 2013.

D. M. Snider and P. J. O'Rourke. The multiphase particle-in-cell (MP-PIC) method for dense particle flow. *Computational Gas-Solids Flows and Reacting Systems: Theory, Methods and Practice* (IGI Global), 2011.

T. Tashimo, T. Suto, J. Murota, and K. Kato. Calcination of fine limestone particles by a powder-particle fluidized bed, *Journal of chemical engineering of Japan*, 32: 374-78, 1999.

L. A. Tokheim, A. Mathisen, L. E. Øi, Ch. Jayarathna, N. H. Eldrup, and T. Gautestad. Combined calcination and CO2 capture in cement clinker production by use of electrical energy. *In TCCS–10. CO2 Capture, Transport and Storage. Trondheim 17th–19th June 2019. Selected papers from the 10th International Trondheim CCS Conference*. SINTEF Academic Press, 2019.

N. Yang, W. Wang, W. Ge, L. Wang, and J. Li. Simulation of heterogeneous structure in a circulating fluidized-bed riser by combining the two-fluid model with the EMMS approach, *Industrial & Engineering Chemistry Research*, 43: 5548-61, 2004.

Comparison of experimental and computational study of the fluid dynamics in fluidized beds with agglomerates

Nora C I S Furuvik   Rajan Jaiswal   Britt M E Moldestad

Department of Process, Energy and Environmental Technology, University of South-Eastern Norway, Norway,
{nora.c.i.furuvik, rajan.jaiswal, britt.moldestad}@usn.no

Abstract

Particle agglomeration is one of the obstacles for successful application and commercial breakthrough of fluidized bed biomass gasification. The problem is generally associated with molten ash components that interact with the bed particles, forming agglomerates that interfere with the flow behavior.

In this work experimental and computational study are combined in order to gain more insight into the fluid dynamics in a bubbling fluidized bed gasifier. The goal is to develop a Computational Particle Fluid Dynamic (CPFD) model that can be used in further investigations of the correlation between flow behavior and bed agglomeration during biomass gasification in fluidized beds. The experimental part was performed in a 20 kW laboratory scale bubbling fluidized bed system. The commercial CPFD software Barracuda Virtual Reactor (VR) version 17.4.1 was used for the computational study. Simulation results were compared to the experimental data in order to validate the CPFD model. Pressure drops predicted by the simulations were in good agreement with the experimental measurements, which indicate that the model is well capable of studying the fluid dynamics in a fluidized bed system.

Keywords: biomass gasification, fluidized bed, agglomeration, CPFD simulations

1 Introduction

Fluidized bed reactors have a broad use in various industrialized applications and are common in both petroleum and petrochemical processes, as well as heat and power production. A typical fluidized bed system consists of a cylindrical column packed with a suitable bed material, which is kept in a fluidized state by passing a fluid through at a velocity that is sufficiently high to "loosen" the bed particles. The fluidized bed design allows for good mixing in all directions within the reactor, resulting in enhanced fuel/fluid contact and thereby increased heat and mass transfer (Sansaniwal, 2017). As a result of the combination of intense solid mixing and bed materials with large thermal capacity, the fluidized beds can be operated under nearly isothermal conditions. Additionally, they have the benefit of continuous and controlled operations (Basu, 2013). Due to their homogenous operation conditions, the fluidized bed reactors are capable of handling a wide range of fuels, and compared to other conversion technologies they are considered well suited for processing highly reactive fuels such as biomass (Basu, 2013; Capareda, 2014).

Despite the many advantages with the fluidized beds, some difficulties are reported related to the gasification process of biomass-derived fuels. Generally, these problems are associated with ash-melting and following agglomeration of bed material. Biomass fuels refer to a broad variety of feedstock, and are characterized as heterogeneous with widely varying chemical and physical properties (Capareda, 2014). Due to the differences in chemical and physical properties, the biomass fuel characteristics are associated with diversity in composition of ash forming elements, which may represent significant barriers for successful fluidized bed gasification processes (Tiffault *et al*, 2018).

Understanding the fluid dynamics in the fluidized bed is essential for maintaining ideal operational conditions for an appropriate fluidized regime. This work is divided into one experimental part and one simulation part. The experimental setup is a 20 kW laboratory scale bubbling fluidized bed gasifier. The laboratory scale model is used to study the fluidized conditions at different gasification temperatures. Additional experiments with a mix of bed material and agglomerates are performed to investigate the dependence of fluidization on particle shape, size and density. For the simulation part, the commercial CPFD software Barracuda VR version 17.4.1 is used for simulations of the flow behavior in a bubbling fluidized bed gasifier. The data and measurements achieved from the fluidization experiments are used to develop and validate a CPFD-model that can be used for further investigations.

2 Particle agglomeration in fluidized beds

Ash related problems are a key concern in gasification of biomass in fluidized beds. The problems are generally related to alkali ash components that interact with the bed particles, forming agglomerates that cause fluid dynamic disturbances in the bed. The agglomerates interfere with the flow behavior, change the fluidized conditions and make further fluidization impossible. During bed agglomeration processes, the solid mixing becomes ineffective because the agglomerates tend to

obstruct the particles movement, resulting in decreased heat transfer and local temperature deviations that in turn create de-fluidized volumes in the bed (Bartles *et al*, 2008). De-fluidization is described as a total collapse of the fluidized bed, and is recognized by a rapidly decreasing bed pressure drop and/or a substantial change in the bed temperature. In the most severe cases, particle agglomeration results in unscheduled shutdowns of the whole installation (Öhman and Nordin, 2000).

Particle agglomeration in fluidized bed biomass gasification is highly coupled to the high temperature chemistry of the biomass ash, i.e. its melting and the consequently appearance of an alkali liquid phase that glue the ash to the surface of the bed particles (Öhman et al., 2000). The mechanisms are dominated by a combination of ash particles attaching the surface of the bed material and chemical reactions that occur between the ash-coated bed particles and the condensed gaseous alkali components (Figure 1).
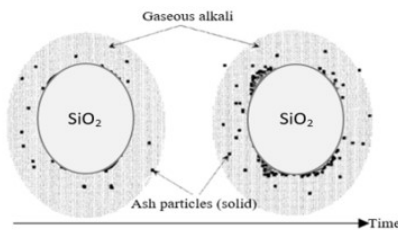
Figure 1. Ash deposition onto the surface of silica sand bed material (Öhman et al., 2000).

As a consequence of repeated collisions between the ash-coated particles in the bed, the particles stick together and eventually they grow towards larger agglomerates. Figure 2 shows a photo of agglomerate formed during gasification experiments with grass pellets in a bubbling fluidized bed gasifier (Furuvik *et al*, 2020).


Figure 2. Agglomerate of biomass ash and silica sand particles, formed in a bubbling fluidized bed gasifier.

3 Experimental setup

The fluidization experiments were performed in a 20 kW laboratory scale bubbling fluidized bed reactor with a height of 100 cm and an inner diameter of 10 cm. A schematic of the experimental setup is shown in Figure 3. The fluidizing agent was air introduced into the bed through two pipes from the bottom of the column. The

air mass flow rate was controlled with a Brook air flowmeter, with an operating range between 0.5 kg/h – 4.7 kg/h. Five pressure transducers placed along the height of the reactor were constantly monitoring the operating conditions in the bed. Each pressure transducer measures the gauge pressure in the given position, i.e. the fluid pressure above the atmospheric pressure. The temperature and pressure sensors are connected to the LabVIEW software for data acquisition. The locations of the pressure and temperature sensors (P1/T1, P2/T2, P3/T3, P4/T4 and P5/T5) are seen in Table 1.
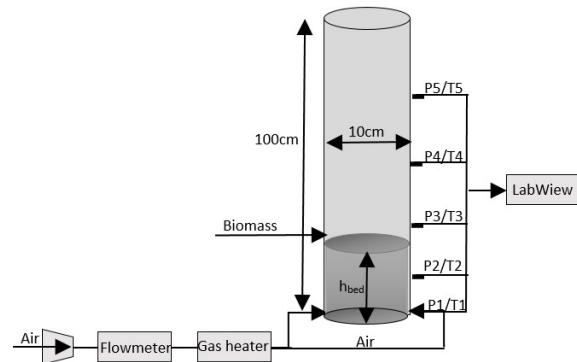
Figure 3. Schematic of the laboratory scaled bubbling fluidized bed used in fluidization experiments.

Silica sand particles with a mean diameter of 305 μm and particle density 2600 kg/m³ were used as bed material. The sand particles were preheated to the operating temperature by letting the air pass through a preheated chamber before entering the reactor. The experiments were run without feeding biomass to the reactor. The externally heated reactor was operated at temperatures that were kept constant throughout each test run. Five thermocouples were used to determine the temperature profile in the bed and to control the temperatures at experimental conditions of 300°C, 600°C, 700°C and 800°C. For each temperature, a fluidization experiment of the bed particles were performed. The pressure drop in the bed were recorded at different superficial air velocities ranging from 0.029 to 0.330 m/s. The superficial air velocities (u_f) were calculated from the mass flow rate ($\dot{m}$), the area of the reactor (A) and the air density at the specific temperatures (ρ_f):

$$u_f = \frac{\dot{m}}{A \cdot \rho_f} \qquad (1)$$

Table 1 lists the operating parameters for the bubbling fluidized bed reactor. The minimum fluidization velocity for each experimental condition was calculated using the equation for minimum fluidization derived from Ergun's equation (Kuuni and Levenspiel, 1991).

Table 1. Operating parameters for fluidization experiments.

Description	Value
Operating temperatures	300, 600, 700 and 800°C
Air flow rate	0.5 – 3.0 kg/h
Superficial air velocity range	0.029 - 0.172 m/s @ 300°C 0.044 - 0.262 m/s @ 600°C 0.049 - 0.292 m/s @ 700°C 0.054 - 0.330 m/s @ 800°C
Calculated minimum fluidization velocity	0.050 m/s @ 300°C 0.038 m/s @ 600°C 0.036 m/s @ 700°C 0.034 m/s @ 800°C
Pressure and temperature measurement locations	P1/T1: 0.023 m P2/T2: 0.143 m P3/T3: 0.238 m P4/T4: 0.538 m P5/T5: 0.838 m

The mass of the bed particles was 2.355 kg, corresponding to a static bed height of 20 cm. For the temperatures of 700°C and 800°C additional test runs were carried out, for these runs a mix of agglomerates of different sizes was introduced to the bed together with the bed materials. The mass of agglomerates was 116 g corresponding to a bed agglomeration of 5% by weight. The agglomerates were produced from previous performed gasification experiments (Furuvik *et al*, 2020). Detailed specification of the properties of the bed material and the agglomerates are shown in Table 2.

Table 2. Specification of particle properties.

Particle properties	Value
Mass of bed material	2.355 kg
Bed material particle size range	180 – 710 μm
Particle density of bed material	2600 kg/m³
Mass of agglomerates	116 g
Size of agglomerates	1.0 – 3.0 cm
Agglomerate density	1510 kg/m³

4 Simulation model

4.1 CPFD model description

CPFD simulations are useful tools in modelling of fluid-particle interaction in fluidized bed reactors. In this study, the commercial software package Barracuda VR was used to simulate the fluid dynamics in a 20 kW laboratory scale bubbling fluidized bed system. Barracuda VR uses the three-dimensional Multiphase Particle-in-Cell (3D-MP-PIC) method for calculating the fluid-particle flow. The method is based on the

Eulerian-Lagrangian approach, wherein the Eulerian approach is used for solving the continuum phase and the Lagrangian approach is used for solving the particle phase. In the MP-PIC method, the solid particles are modeled as computational particles with a proper size and density distribution (Jayarathna *et al*, 2017; Thapa and Halvorsen, 2014). The governing equations include the conservation of mass, momentum and energy in the system. The interphase momentum transfer is an important term when modelling fluidized bed systems, and is described in details by Chladek *et al*. (Chladek *et al*, 2018) and Jayarathna et al. (Jayarathna *et al*, 2017).

The particle fluidization results from the drag forces exerted by the fluid on the particles. The drag forces are the main cause of transfer of mass, momentum and energy between the different phases in the fluidized bed system (Marchelli *et al*, 2020). In Barracuda VR the drag forces (F) have their general form:

$$F = m_p D(u_f - u_p) \qquad (2)$$

Where m_p is the mass of the particles, u_f is the fluid velocity, u_p is the particle velocity and D is the drag function (CPFD Software, 2020). The dimensionless drag function is the fluid-particle interphase exchange coefficient and differs for the different drag models. Common for all systems are that D always has a complex dependency on the bed porosity and the particle Reynolds number (Re) (Marchelli *et al*, 2020). The Re is defined as:

$$Re = \frac{2r_p \rho_f (u_f - u_p)}{\mu_f} \qquad (3)$$

Where r_p is the particle radius, ρ_f is the fluid density and μ_f is the fluid viscosity. The drag models determine the drag forces acting on the particles, and several drag models are available in Barracuda. In order to study the behavior of different drag models for the chosen system, the Wen-Yu drag model (CPFD Software, 2020; Wen and Yu, 1966), the Ergun drag model (CPFD Software, 2020; Ergun, 1952) and the Wen-Yu/Ergun drag model (CPFD Software, 2020) were tested.

The Wen-Yu model is considered most valid for dilute systems. The drag function for the Wen-Yu model (D_{WY}) is dependent on the fluid conditions and the particle properties, and is related to the drag coefficient (C_d) (CPFD Software, 2020; Wen and Yu, 1966):

$$D_{WY} = \frac{3}{8} C_d \frac{\rho_f (u_f - u_p)}{\rho_p r_p} \qquad (4)$$

Where ρ_p is the particle density.

The drag coefficient (C_d) is defined as a function of Re and is calculated by the following set of equations (CPFD Software, 2020; Wen and Yu, 1966):

$$C_d = \begin{cases} \dfrac{24}{Re}\theta_f^{n_0} & Re < 0.5 \\[2mm] \dfrac{24}{Re}\theta_f^{n_0}(c_0 + c_1 Re^{n_1}) & Re \leq 0.5 \leq 10000 \\[2mm] c_2\theta_f^{n_0} & Re > 10000 \end{cases} \quad (5)$$

Where θ_f is the fluid volume fraction, c_0, c_1, c_2, n_0 and n_1 are model constants with default values 1.0, 0.15, 0.44, -2.65 and 0.687 respectively.

The Ergun drag model is developed from dense bed data and is primarily most suitable for picturing flow through static packed beds. The drag function (D_E) is given by (CPFD Software, 2020; Ergun, 1952):

$$D_E = 0.5\left(\frac{k_1\theta_p}{\theta_f Re} + k_0\right)\frac{\rho_f(u_f - u_p)}{\rho_p r_p} \quad (6)$$

Where θ_p is the particle volume fraction, k_0 and k_1 are constants with default values 2 and 180 respectively.

Wen-Yu/Ergun drag model is a combination of the Wen-Yu and the Ergun models. This allows it to be able to work well in both dense and dilute systems. The drag function (D_{WYE}) is controlled by the close pack volume fraction (θ_{cp}) with a switch from Ergun to Wen-Yu at defined values. Wen-Yu/Ergun uses the Ergun function for $\theta_p > 0.85\cdot\theta_{cp}$ and the Wen-Yu function at higher voidage (CPFD Software, 2020).

$$D_{WYE} = \begin{cases} D_{WY} & \theta_p < 0.75\theta_{cp} \\[2mm] D_{E\_WY} & 0.75\theta_{cp} \geq \theta_p \geq 0.85\theta_{cp} \\[2mm] D_E & \theta_p > 0.85\theta_{cp} \end{cases} \quad (7)$$

Where D_{E_WY} is defined as:

$$D_{E\_WY} = (D_E - D_{WY})\left(\frac{\theta_p - 0.75\theta_{cp}}{0.85\theta_{cp} - 0.75\theta_{cp}}\right) + D_{WY} \quad (8)$$

4.2 CPFD simulations

CPFD simulations of a 20 kW laboratory scale fluidized bed were performed. The 3D computational grid was created using 10000 control volumes. The reactor was initially loaded with silica sand particles with mean particle diameter of 305 μm and a particle density of 2600 kg/m³. The particle size distribution is determined based on the discrete mass frequency distribution (Crowe *et al*, 2012), the result is shown in Figure 4.
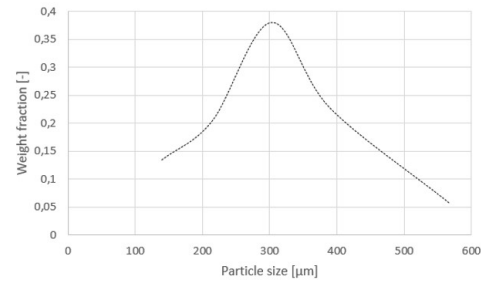
Figure 4. Particle size distribution of bed material.

Fluidizing agent was air at atmospheric pressure. Flow boundary conditions were applied at the bottom of the reactor. The pressure in the bed was measured at positions P1= 0.023 m, P2 = 0.103 m and P3 = 0.183 m above the bottom of the column. The simulations were run for 70 s with a time step of 0.0001 s. Table 3 shows the values of the model parameters used in the simulations. In order to add agglomerates to the fluidized bed a coarser grid was required, and the number of grid cells was therefore reduced from 10000 to 5120. The size of the agglomerates was limited by the chosen grid, which allowed a maximum particle size of 1.0 cm. In the present simulations, the size of the agglomerates ranged from 0.5 cm to 1.0 cm with a particle density approximately equal to 1510 kg/m³ (Furuvik *et al*, 2019).

Table 3. Model parameters used in the CPFD simulations.

Description	Value
Particle density	2600 kg/m³
Fluidizing agent	Air
Type of flow	Isothermal@ 300, 600, 700 and 800°C
Particle size	Range: 180 - 710 μm Mean diameter: 305 μm
Close-pack volume fraction	0.6
Particle sphericity	0.86
Static bed height	0.20 m
Superficial gas velocity	0.005 – 0.200 m/s
Agglomerate particle size	1.0 cm
Agglomerate density	1510 kg/m³

5 Results and discussion

The bed pressure drop was measured experimentally at different superficial gas velocities. The operating temperatures were 300°C, 600°C, 700°C and 800°C. Figure 5 shows that the pressure drops decrease with increasing bed temperatures. The drag equations explain

how the bed operating temperature alters the fluidized conditions in the bed. All drag functions indicate that the drag forces are strongly dependent on both the bed porosity and the Re. Moreover, increased bed temperature results in increased fluid viscosity (μ_f) and decreased fluid density (ρ_f), and hence lower Re. From the Wen-Yu drag functions, equation (3) and (4), it is obvious that changing the Re will cause a change in the magnitude of drag forces exerted on the bed particles. Stronger drag forces acting on the bed material give lower pressure drop in the fluidized bed.
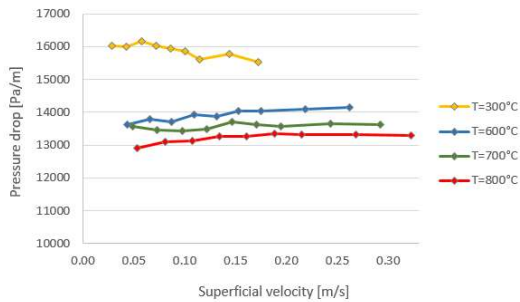
Figure 5. Comparison of experimental bed pressure drops with different operating temperatures.

Additional two fluidization experiments were carried out, where 5 wt% of agglomerates were mixed together with the bed particles. The experimental temperature conditions were 700°C and 800°C. Figure 6 shows that adding agglomerates to the process alters the character of the fluidized conditions. This can also be seen in drag force calculations using the drag functions described in equation (3), (5) and (7). The drag functions describe how the relation between the superficial velocity and the particle shape, size and density determine the bed conditions during fluidization.

The agglomerates are relatively big, but porous, which give them low particle density (ρ_p). A change in the ρ_p alters the gravitational forces acting on the particles. Looking at the drag equations, lower ρ_p gives increased drag function that further results in stronger drag forces and decreased pressure drop.

The agglomerates are more angular compared to the sand particles. Lower sphericity alters the packing properties of the bed and leads to change in the associated void spaces. Larger voids in the bed result in higher fluid volume fraction, which based on the Wen-Yu drag coefficient give increased drag forces and thereby lower bed pressure drop.
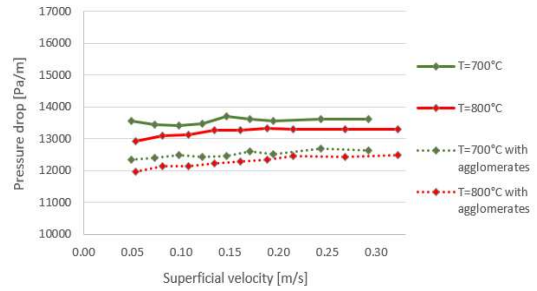
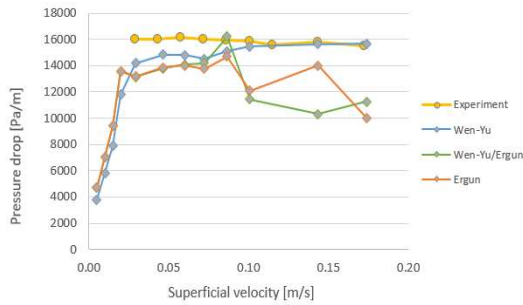**Figure 6.** Comparison of experimental bed pressure drops with and without agglomerates.

To predict the fluidized conditions in the bubbling fluidized bed, the experimental setup was modeled using CPFD calculations. The simulation model was validated by comparing the measured pressure drop from experiments with results from the CPFD simulations. Wen-Yu, Ergun and the combined Wen-Yu/Ergun drag models were tested in order to study the behavior of the different drag models for the present system. Figure 7 compares the experimental results and the results from the simulations of the three different CPFD models at bed temperature of 300°C. The results show that the Wen-Yu drag function gives better prediction compared to the other drag models.

The advantage of Wen-Yu is that the drag force only depends on the fluid volume fraction and the Re, which make it very suitable for predicting the fluid dynamics in stabilized systems with isothermal bed conditions. The Wen-Yu drag function is based on a dependence on the Re, with a switch between different functions for Re<0.5 and Re>1000. Re increases as the superficial fluid velocity increases. However, Re will never exceed 1000 nor fall below 0.5 in the selected superficial velocity range. For this system, Wen-Yu therefore uses the same equation to calculate the bed conditions for all measuring points during the simulations. As the bed temperature is kept constant during the whole simulation time, it can be assumed that both the fluid conditions and the particle properties are unchanged. Fluid volume fraction will admittedly fluctuate slightly as a result of where and how bubbles are formed in the bed. These fluctuations are relatively small, hence they will not give large disturbances in the fluidized conditions.

Ergun drag model is based on data from fixed bed experiments and is therefore expected to be more appropriate at higher packing fractions. The superficial fluid velocity has large contribution to Ergun equation. For low velocities, the bed conditions are mainly controlled by the particle packing. As the superficial velocity increases, the velocity takes more control over the bed conditions. At higher velocities, Ergun gives large fluctuations in the pressure drop. As seen in Figure 7, the drag model fails in the fluidized regime.

Wen-Yu/Ergun drag model is a combination of the Wen-Yu model and the Ergun model, whereas the close-

pack volume fraction (θ_{cp}) determines which drag function that are used. For this system with $\theta_{cp} = 0.6$, the Ergun equation is applied when the particle volume fraction (θ_p) > 0.51, and Wen-Yu is applied when θ_p < 0.45. The results indicate that the only drag function used for the present CPFD calculations is Ergun, which explains why the simulated pressure drop for Ergun and Wen-Yu/Ergun drag models are about the same.

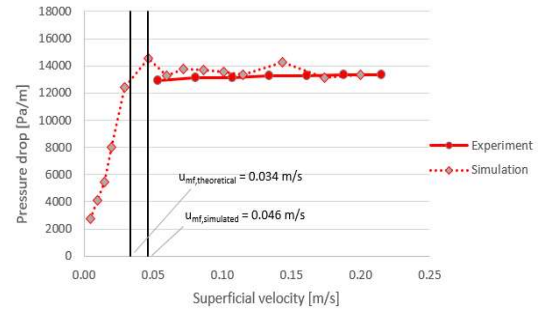**Figure 7.** Comparison of experimental result and simulations using different drag models.

Figure 8 shows the experimental and the simulated pressure drops in the fluidized bed for temperature conditions of 300°C. The theoretical minimum fluidization velocity (u_{mf}) for the particles was previously calculated to $u_{mf,theoretical} = 0.05$ m/s. From the CPDF simulations, the minimum fluidization velocity is found at approximately the same value, $u_{mf,simulated} = 0.05$ m/s. Comparison of the experimental pressure drop and the simulations using the Wen-Yu drag model shows that the CPFD model can predict fluid dynamic behavior of fluidized bed reasonably well.
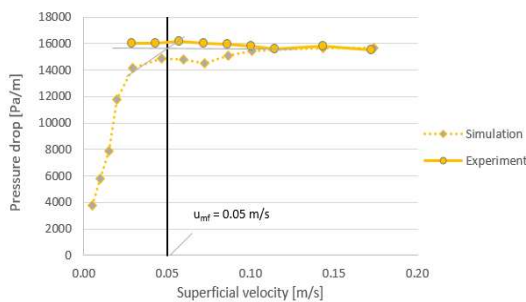
Figure 8. Comparison of experimental and simulated bed pressure drops for 300°C.

Figure 9 shows the experimental and the simulated pressure drops in the fluidized bed for temperature conditions of 800°C. As seen in the figure, the minimum fluidization velocities have decreased with the increased temperature. u_{mf} is indicated by black vertical lines in the figure, which read off $u_{mf,theoretical} = 0.034$ m/s and $u_{mf,simulated} = 0.046$ m/s. The large deviation between $u_{mf,theoretical}$ and $u_{mf,simulated}$ can be explained by the theoretical calculation using the mean particle diameter of 305 μm, while the CPFD calculation use the particle size distribution (Figure 4) where the particle diameter ranges from 180-710 um. However, the CPFD

simulation correctly predicts the fluidized regime and pressure drops in the fluidized bed system.

Figure 9. Comparison of experimental and simulated bed pressure drop at 800°C.

Figure 10 and 11 compare the experimental and the simulated pressure drops for fluidization of sand and sand mixed with agglomerates for the temperature conditions of 700°C and 800°C respectively. One problem in the application of CPFD modelling of the agglomerated fluidized bed systems is that the agglomerates exist in all types of size, shapes and structures, which makes it difficult to define the agglomerated particles properties correctly. Although the simulations show instabilities in the fluidized regimes, the CPFD model maintains good agreement for the fluidized operation conditions in the fluidized regime.
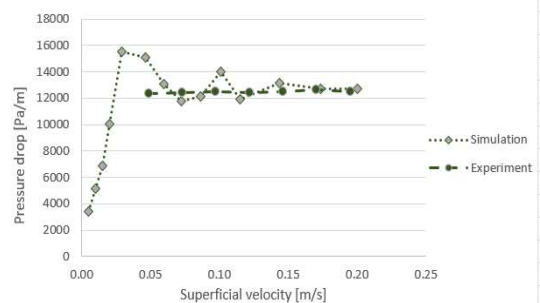
Figure 10. Comparison of experimental and simulated pressure drop for agglomerated bed at 700°C.
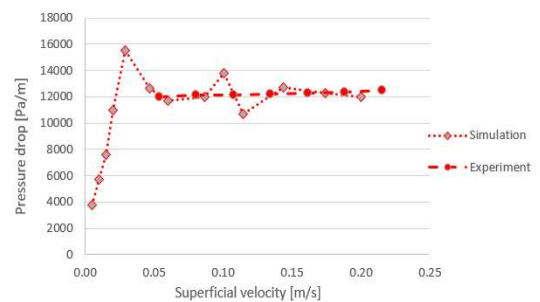
Figure 11. Comparison of experimental and simulated pressure drop for agglomerated bed at 800°C.

6 Conclusion

In this study, the fluid dynamics of a bubbling fluidized bed gasification system were investigated computationally and experimentally. The study included CPFD simulations of fluidization of silica sand particles and agglomerates using the commercial simulation software Barracuda VR version 17.4.1. Comparison of experimental and simulated pressure drops over the bed showed that the model can predict fluid dynamic behavior of fluidized bed reasonably well. Furthermore, the comparison showed that the Wen-Yu drag model gave better prediction of the fluidized conditions in the bed compared to the Ergun and the Wen-Yu/Ergun drag models.

The agglomerates are large sized and porous, which give them low density. The fluid dynamics in the bed depend upon the particle shape, size, density and diameter. Thus, the agglomeration process disturbs the fluidized conditions in the bed. The CPFD model is well capable of predicting the effect of agglomerates on flow behavior in a fluidized bed gasifier, and can be used for further studies including ash from different types of biomass.

Acknowledgements

References

M. Bartles, W. Lin, J. Nijenhuis, F. Kapteijn and R. Ommen. *Agglomeration in fluidized beds at high temperaures: Mechanisms, detection and prevention*. Progress in Energy and Combustion Science 34:633-666, 2008.

P. Basu. *Biomass Gasification, Pyrolysis and Torrefaction*, Second Edition. Academic Pess, UK, 2013.

S. C. Capareda. *Introduction to Biomass Energy Conversions*. CRC Press, Boca Raton, US, 2014. ISBN: 978-1-4665-1334-1 (eBook-PDF).

J. Chladek, C. K. Jayarathna, B. M. E. Moldestad and L- A. Tokheim. *Fluidized bed classification of particles of different size and density*. Chemical Engineering Science 177: 155-162, 2018.

CPFD Software, LLC. *Barracuda VR Solutions,* (2020). Available from: https://cpfd-software.com/barracuda-vr-solutions/barracuda-vr (Accessed July, 2020)

C. T. Crowe, J. D. Schwarzkopf, M. Sommerfield and Y. Tsuji. *Multiphase Flows with Droplets and Particles*, Second Edition. CRC-Press, USA, 2012.

S. Ergun. *Fluid flow through packed columns*. Chemical Engineering Progress, 48: 89, 1952.

N. C. I. S. Furuvik, R. Jaiswal and B.M.E. Moldestad. *Experimental study of agglomeration of grass pellets in fluidized bed gasification*. WIT Transactions on Ecology and the Environment, 246, 2020.

N. C. I. S. Furuvik, R. Jaiswal and B. M. E. Moldestad. *Flow behavior in an agglomerated fluidized bed gasifier*. International journal of Energy and Environment 10(2):55-64, 2019.

C. K. Jayarathna, B. M. E. Moldestad and L. A. Tokheim. *Validation of results from Barracuda® CFD modelling to predict minimum fluidization velocity and pressure drop of Geldart A particles*. Proceedings of the 58th SIMS conference, 2017.

D. Kunii and O. Levenspiel. *Fluidization Engineering*, Second Edition. Butterworth-Heinemann, USA, 1994

F. Marchelli, Q. Hou, B. Bosio, E. Arato and A. Yu. *Comparison of different drag models in CFD-DEM simulations of spouted beds*. Powder Technology 360:1253-1270, 2020.

S. K. Sansaniwal, K. Pal, M. A. Rosen and S. K. Tyagi. *Global challenges in the sustainable development of biomass gasification: An overview*. Renewable and Sustainable Energy Reviews 80: 23-43, 2017.

R. K. Thapa and B. M. Halvorsen. *Study of Flow Behavior in Bubbling Fluidized Bed Biomass Gasification Reactor using CFD simulation*. The 14th International Conference on Fluidization - From Fundamentals to Products, Eds, ECI Symposium Series, 2013.

E. Tiffault, S. Sokhansanj, M. Ebadian, H. Rezaei, E. O. B. Ghiasi, F. Yazdanpanah, A. Asikainen and J. Routa. *Biomass pre-treatment for bioenergy. Case study 2: Moisture, physical property, ash and density management as pre-treatment practices in Canadian forest biomass supply chains*. IEA Bioenergy, 2018.

C. Wen and Y. Yu. *Mechanics of fluidization*. Chemical Engineering Progress Symposium Series, 62: 100-111, 1966.

L. C. Williams, T. L. Westover, R. M. Emerson, J. S. Tumuluru and C. Li. *Sources of Biomass Feedstock Variability and the Potential Impact on Biofuels Production*. Bioenergy Resource 9: 1-14, 2016.

M. Öhman, A. Nordin, B.J Skrifvars, R. Backman and M. Hupa. *Bed Agglomeration Characteristics during Fluidized Bed Combustion of Biomass Fuels*. Energy & Fuels 14(1): 169-178, 2000. DOI: 10.1021/ef990107b

M. Öhman and A. Nordin. *The Role of Kaolin in Prevention of Bed Agglomeration during Fluidized Bed Combustion of Biomass Fuels*. Energy & Fuels 14(3): 618-624, 2000. DOI: 10.1021/ef990198c

Computational modeling of fluidized bed behavior with agglomerates

Krister Jakobsen   Rajan Jaiswal   Nora C. I. S. Furuvik   Britt M. E. Moldestad

Department of Process, Energy and Environmental Technology, University of South-Eastern Norway, Norway
Krister.Jako@gmail.com {Rajan.Jaiswal, Nora.C.I.Furuvik, Britt.Moldestad}@usn.no

Abstract

Fluidized bed reactors can be used for biomass gasification. The product from biomass gasification is syngas, which can be used for production of bio oil. The main challenge when using fluidized bed for gasification is ash melting and agglomeration of the bed material. Agglomeration of the bed material influences on the flow behavior in the fluidized bed reactor and thus affects the gasification efficiency. A Computational Particle Fluid Dynamic (CPFD) model is developed to predict the flow behavior in a fluidized bed gasifier. The CPFD model was validated against experimental data from a cold fluidized bed. The model was then tested against the results from a biomass gasifier, and a few modifications were needed. Glickman's scaling parameters were used to scale up from a lab-scale to a full-scale gasifier. Simulations using the modified model were performed to study the flow behavior in a full-scale gasifier with agglomerates. It was found that the CPFD model is capable of predicting the effect of agglomerates on flow behavior in a fluidized bed gasifier.

Keywords: biomass, gasification, ash, agglomeration, CPFD, Barracuda

1 Introduction

Biomass is considered a renewable energy source, and it is crucial to make the biomass conversion processes more energy effective. Biomass is converted via gasification into a syngas consisting of mainly CO and H_2. Different technologies are used for gasification of biomass, and one of the most promising technologies is fluidized bed reactor. Fluidized bed reactors are used to ensure proper mixing of biomass and fluidizing gas, and thus increase the heat and mass transfer. Fluidized bed reactors are also very flexible when it comes to the type and quality of the biomass feed. The challenges when using fluidized bed gasifiers are ash melting and agglomeration of the bed material. Agglomeration may disrupt the flow behavior in the fluidized bed and thus affect the overall efficiency of the gasifier (Basu, 2013).

2 Ash melting and agglomerates

Ash melting is a big challenge in operation of biomass fluidized bed gasifiers. The amount of ash from biomass varies a lot depending on the type of biomass used. The typical content of ash in wood chips, straw and solid municipal waste is about 1 wt%, 8 wt% and 50 wt% respectively. However, even a small amount of ash can harm the gasification process, and it is therefore important to study how ash melting and agglomeration affect the flow behavior in a fluidized bed. Biomass such as grasses, demolition wood, and straw have a high potential to create agglomeration, fouling, and corrosion in a gasifier. (Basu, 2013)

The operating temperature for biomass fluidized bed gasifiers is usually kept in the range of 700-900°C to avoid ash melting. Ideally, the temperature should be increased to obtain a higher quality of the syngas and to avoid problems with tar. The ash melting temperature varies depending on the composition of the biomass. The melting temperature for spruce wood is 1170°C, whereas the melting point for wheat straw is 915°C (Basu, 2013).

The most significant ash-forming elements in biomass are silica (Si), potassium (K), calcium (Ca), magnesium (Mg), aluminum (Al), phosphor (P), chlorine (Cl), sodium (Na) and sulphur (S). (Balland et al., 2017; Furuvik et al., 2020). The ash-forming elements are released from the biomass during the heating process. When the ash melts, the inorganic elements from the melted ash can create a sticky component, which functions as a glue between the ash and the sand particles, and agglomerates are formed. Figure 1 shows agglomerates which are created by melted biomass ash and sand particles. (Furuvik et al., 2019a; Furuvik et al., 2019b). The agglomerates are of various sizes and shapes, and may change the fluidization properties in a biomass gasifier significantly. An agglomerated bed creates instability in the bubble frequency and can cause fluid channeling. Agglomeration can also result in defluidized zones in the gasifier. In the most severe cases, particle agglomeration may lead to unscheduled shutdowns of the whole installation. (Öhman et al., 2000)

**Figure 1.** Agglomerates formed during gasification of biomass.

3 Material and methods

3.1 Experimental set-up

Two different experimental set-ups have been used in this study. One is a cold fluidized bed, where the fluid dynamic properties and flow behavior are studied. The other one is a biomass gasification reactor used to study the gasification yield at different operation conditions.

3.1.1 Cold fluidized bed set-up

The cold fluidized bed consists of a transparent cylindrical tube open to the atmosphere, and with a gas distributor plate at the bottom. The height and diameter of the cylinder are 140 cm and 8.4 cm, respectively. Pressure transducers are installed along the height of the bubbling fluidized bed, and the distance between the transducers is 10 cm. The model is shown in Figure 2. Experiments were performed with sand particles with density 2650 kg/m^3 as the bed material. A sieving analysis of the sand particles was performed and it was found that the particles had a size range of 300-700 µm. The weighted mean particle diameter was calculated to 535 µm. An aspect ratio (bed height/bed diameter) of 2.5 was used in the experiments. Pressure drop in the bed was monitored and plotted versus the superficial air velocity.
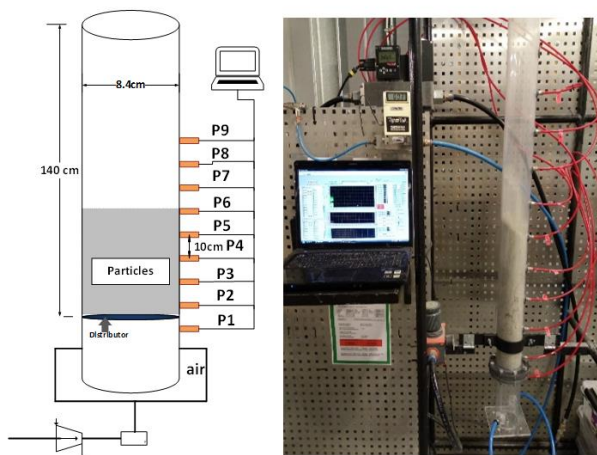
Figure 2. Experimental set-up; cold fluidized bed (Jaiswal *et al*, 2018).

3.1.2 Biomass gasifier set-up.

A lab-scale bubbling fluidized bed reactor with a feed capacity of 3–5 kg/h was used for the experimental biomass gasification tests. A drawing of the reactor is presented in Figure 3. The experimental rig is composed of a feeding system, a fluidized-bed reactor, a pre-heating system for the fluidizing gas stream, a sampling line and an exhaust line. The biomass gasification rig is made of stainless steel. It has three electrical heating elements, which are installed externally. The gasifier is insulated with refractory material on the inside, and a 200 mm thick fiberglass layer on the outside to minimize the heat losses.

The feeding system consists of a silo followed by a cold and a hot screw feeder. The cold screw feeder conveys the feed from the silo to the hot screw feeder. The hot screw feeder transports the feed into the reactor. The reactor is a cylindrical vessel with an inner diameter of 0.1 m and a height of 1.0 m. Air is used as the fluidizing agent, and flows through a pre-heater with a capacity of 18 kW. The exhaust line goes from the top of the reactor and transports the product gases into a flare. At the top of the reactor, there is also a sampling line, where samples of the syngas are taken at regular intervals and analyzed in a gas chromatograph (GC SRI 8610C). Two experiments were performed at temperature 735°C using gas flow rates of 1.5 and 2 kg/h. The bed material was sand with a mean particle diameter of 367 µm and density 2650 kg/m^3. The pressure drop over the particle bed was monitored.
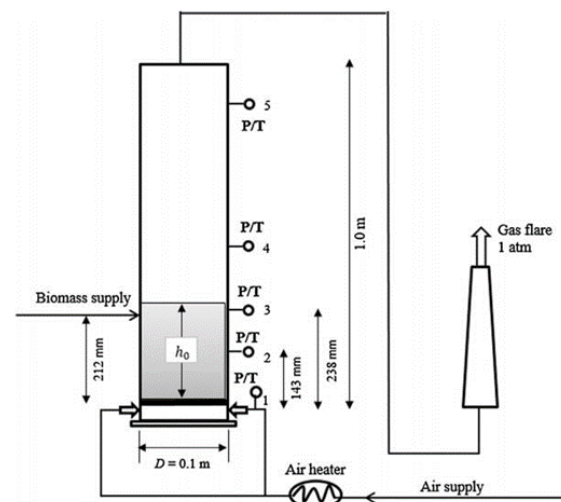
Figure 3. Experimental set-up; biomass gasifier.

3.2 CPFD modelling

Computational Particle Fluid Dynamics (CPFD) simulations are used to predict the minimum fluidization velocity and the flow behaviour in a cold fluidized bed and in a gasification reactor. The bed dimensions and the fluid and particle properties were the same as in the experimental tests. The CPFD model is developed using Barracuda VR. The CPFD numerical methodology

incorporates the multi-phase-particle-in-cell (MP-PIC) method, where particles with the same properties are grouped into parcels and each parcel is represented by one computational particle (Snider, 2001; Amarasinghe, 2017). The gas phase and the particle phase are modelled using the Eulerian and the Lagrangian approaches respectively. Chladek et al. (2018) and Jayarathna et al. (2017) have described the transport equations in detail. Several drag models are available in Barracuda, and five of them were tested in this study. The five drag models are Ergun, Wen-Yu, Wen-Yu&Ergun, Turton-Levenspiel and Nonspherical Haider-Levenspiel. The Ergun drag model is derived based on dense bed systems, and is only valid for gas volume fractions lower than 0.8. The Wen–Yu drag model (Wen and Yu, 1966) is developed based on a series of experiments performed by Richardson and Zaki in 1954. The experimental data are available in (Richardson and Zaki, 1997). The Wen-Yu correlation is valid when the internal forces between particles are negligible, meaning that the viscous drag forces dominate the flow behavior. The Wen-Yu drag coefficient is a function of the Reynolds number and the void fraction. Researchers have reported good agreement between experiments and simulations using the Wen-Yu drag model (Furuvik et al., 2019). The Wen-Yu & Ergun drag model, is a combination of the two drag models. Wen-Yu's drag model is for dilute systems, and the Ergun drag model is for dense systems. This blend of drag models is controlled by the conditions set by the particle volume fraction and close pack volume fraction. Bandara et al. obtained good results when using the Wen-Yu & Ergun drag model in simulation of a circulating fluidized bed system (Bandara et al., 2019). The Turton-Levenspiel model (Turton and Levenspiel, 1986) and the non-spherical Haider-Levenspiel model (Haider and Levenspiel, 1989; Chhabra et al., 1999) utilize a single particle drag function dependent on the fluid volume fraction.

3.3 Scaling

Fluidized bed reactors are operating under relatively high temperatures, and it is difficult to observe the flow behavior during operation. Therefore, cold fluidized beds are often used for these types of studies. A CFD or CPFD model can be developed and validated against experimental data from cold bed tests. The model can further be used for a biomass gasifier operating at high temperatures. The cold bed has to be scaled based on scaling rules to get the correct dimensions for the particles and reactor, and to fit with the flow behavior observed in the cold fluidized bed. Scaling rules are used to scale from cold to hot, and also for up-scaling from lab scale to pilot or industrial sized reactors. To obtain the fluid dynamic similarities between two reactors, properly developed scaling rules must be used.

The most commonly used scaling rules for fluidized bed reactors are the rules proposed by Glicksman (Glicksman, 1984; Glicksman study, a simplified set of Glicks parameters is used. This simplified set is known as Glicksman's viscous limit set of dimensionless parameters, and can be used when Reynolds number is less than 4. The dimensionless numbers are: $\frac{}{U_0^2}, \frac{L}{d_p}, \frac{L_1}{L_2}, \frac{U_0}{U_{mf}}, \varphi$
$g \cdot L$

where U_0 is the operating gas velocity, minimum fluidization velocity, diameter, L is a reactor dimension (eg. diameter, height, bed height) and φ is the sphericity.

4 Results and discussion

4.1 Validation of CPFD model

A CPFD model of the cold fluidized bed was developed and validated against experimental data. was to find the drag model that gave the best fit experimental data. The grid resolution was set to cells, which resulted in a uniformly distributed grid of 9536 cells. Figure 4 shows the simulated pressure drop versus the superficial velocity for the different drag models.

The Turton-Levenspiel model fits well with the experimental results in the fixed bed area (velocities lower than 0.17 m/s), whereas the Wen a good agreement with the experimental data in the fluidized regime (velocity between 0.16 and 0.20 m/s). A fluidized bed reactor will operate in the fluidized regime, and therefore it was decided to use the Wen drag function for the further validation of the CPFD model.
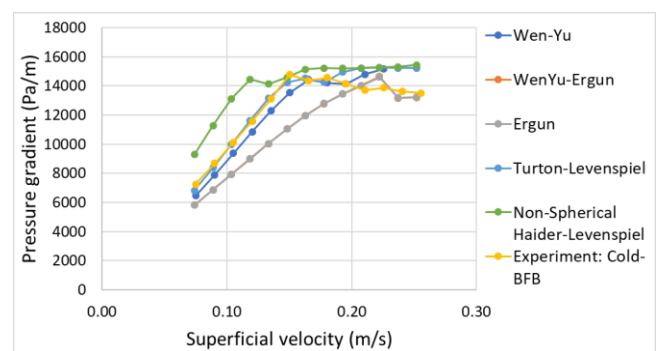
Figure 4. Comparison of different drag models against experimental data.

The next was to perform a grid resolution test. The number of cells was reduced from 12000 to 4000, and it was clear that using 4000 cells gave a significant deviation from the experimental results. was also run with 20000 cells, but this did not give any significant difference compared to using 12000 cells. Since simulation with 20000 cells are more time

consuming, the further simulations were run with 12000 cells. A time step dependency test was carried out, The time steps were changed from 0.001 s to 0.0001 s. The results are presented in Figure 5. The plot shows that the CPFD model using Wen-Yu drag model, 12000 cells and time step 0.0001 s gives a good agreement with the experimental data. This model is further used to simulate agglomerates in an up-scaled fluidized bed reactor.

Figure 5. Comparison of simulations with different time step.

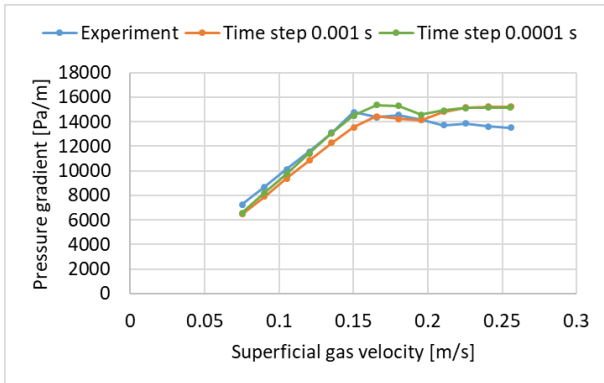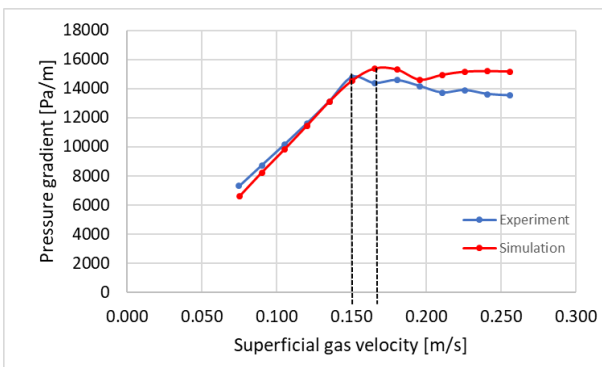Figure 6 shows a comparison of the minimum fluidization velocity from the experiments and the simulation using the final model. The simulated minimum fluidization velocity is about 10% higher than the experimental. The deviation may be due to small differences in particle size distribution and the closed packed volume fractions. The drag model could have been tuned to fit the experimental data better. However, the model is capable of predicting the flow behavior in a fluidized bed with acceptable accuracy.

Figure 6. Comparison of minimum fluidization velocity. Dotted lines mark the minimum fluidization velocities.

The CPFD model was further tested against the results from the biomass gasifier, and it gave a good fit regarding the pressure drop over the bed at gas velocity 0.20 m/s.

4.2 Scaled model

The lab scale fluidized bed gasifier was up-scaled using Glicksman's viscous limit set of dimensionless parameters. Before doing the scaling, the diameter ratio of the two beds was set to 1:5, which means that the up-scaled gasifier has a diameter of 0.5 m. The reason why the geometry ratio was fixed, was to be able to simulate large agglomerates by using the same number of cells as in the simulation of the lab scale gasifier. In CPFD simulations of gas particle systems, the particle sizes cannot exceed the size of the computational cells. The scaling parameter L/d_p had to be neglected to avoid unreasonably large sand particles to be used in the up-scaled bed. The scaling parameters $\frac{U_0^2}{g \cdot L}, \frac{L_1}{L_2}, \frac{U_0}{U_{mf}}, \varphi$, were used for the up-scaling. The calculated parameters for the scaled and the lab scale gasifier are presented in Table 1. The highlighted values in the table are calculated based on the scaling rules.

Table 1: Parameters for the lab-scale and the up-scaled gasifier.

Parameters	Lab-scale	Up-scaled
D	0.1 m	0.5 m
H_{bed}	0.21 m	**1.05 m**
T	735°C	735°C
ρ_p	2650 kg/m^3	2650 kg/m^3
ρ_g	0.35 kg/m^3	0.35 kg/m^3
μ_g	0.000042Pa·s	0.000042Pa·s
d_p	367 μm	**549 μm**
U_0	0.153 m/s	**0.342 m/s**
U_{mf}	0.051 m/s	**0.114 m/s**
d_{aggl}	-	3-4 cm
ρ_{aggl}	1506 kg/m^3	1506 kg/m^3

Simulation with pure sand particles and with sand particles together with agglomerates were performed. The operating air velocity was set constant to 0.15 m/s in both simulations. In the simulation with agglomerates, agglomerates were fed continuously to the gasifier at a mass flow rate of 1.0 kg/s. The simulations were run for 90 s, and the pressure drop over the bed versus superficial gas velocity was monitored. The aim of these simulations was to find the amount of agglomerates required to influence on the flow behavior in the bed. Figure 7 shows the comparison of the simulations with and without agglomerates. As can be seen from the figure, there is a drop in the pressure for the agglomerated bed already after 20 s (20 kg of agglomerates fed to the reactor). The reason can be that the average density of the bed decreases due to the presence of agglomerates having a lower density than the sand. The lower pressure drop can also be due to air channeling caused by the large agglomerates. Between 45 and 70 s the pressure drop over the agglomerated bed decreases gradually from 13500 Pa/m to 12300 Pa/m. During this time interval, the mass of agglomerates in the bed has increased from 45 kg to 70 kg, and it seems that the agglomerates have started to affect the flow behavior significantly. A higher concentration of agglomerates may lead to higher degree of channeling

of the air, which again gives a lower pressure drop over the bed. The problem is that a large part of the air leaves the bed through channels, and that the remaining air is not sufficient to fluidize the bed. This can result in defluidized zones in the bed, low mass and heat transfer and uneven temperature in the bed. Zones with very high temperature can occur, which promotes ash melting and formation of more and bigger agglomerates. In the end, the bed will collapse, and the gasification process has to be shut down. Ash melting and agglomeration of bed material can also plug the reactor totally, and thereby lead to a dangerous situation. It is therefore crucial to ensure that the gasifier is always running in the fluidized regime.
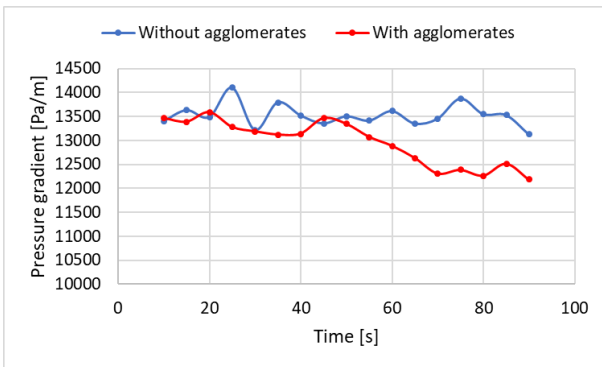
Figure 7. Simulation of gasifier with and without agglomerates. Gas velocity 0.15 m/s. Agglomerate feeding 1 kg/s.

Figure 8 shows the distribution of agglomerates (red) in the gasifier at time 0 s (no agglomerates), 60 s (60 kg of agglomerates and 90 s (90 kg of agglomerates). The agglomerates seems to segregate towards the bottom of the bed as the fraction of agglomerates increases. Figure 9 shows the flow behavior in the bed with 14 kg, 60 kg and 90 kg of agglomerates. The shape and frequency of the bubbles (areas with low particle fractions) change as the fraction of agglomerates increases. Also, the bubbles are getting more diffuse, meaning that the fraction of particles in the bubbles is increasing. This indicates that most of the air is leaving the bed in channels or through the emulsion.
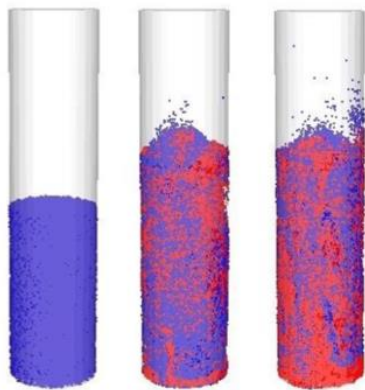
Figure 8. Distribution of agglomerates in the bed at time 0 s, 60 s and 90 s. Sand is blue, agglomerates are red.
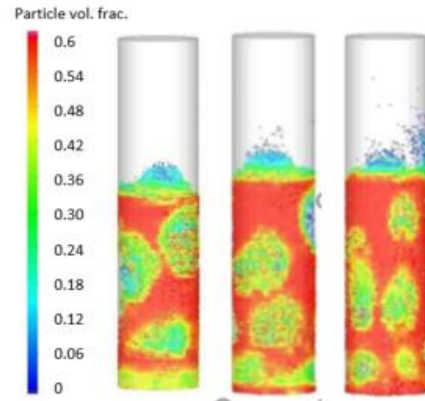
Figure 9. Flow behavior in agglomerated fluidized bed. Gas velocity 0.15 m/s, agglomerates 14, 60 and 90 kg.

The next simulations were run to predict the minimum fluidization velocity for pure sand and for sand with 60 kg (26% by volume) of agglomerates. The results are presented in Figure 10.
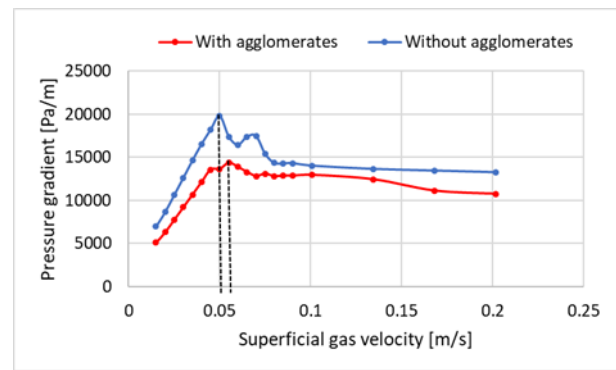
Figure 10. Minimum fluidization velocities (marked with dotted lines) for sand bed and agglomerated bed.

The pressure drop over the bed is significantly lower for the agglomerated bed than for the sand bed both in the fixed bed regime and in the fluidized regime. The minimum fluidization velocities are 0.05 m/s for the sand bed and 0.055 m/s for the agglomerated bed, which is significantly lower than the operating velocity of 0.15 m/s in the previous simulations. It is also observed that the minimum fluidization velocity for the sand bed (0.05 m/s) is less than half of the scaled minimum fluidization velocity (0.114 m/s). The reason is that the scaled minimum fluidization velocity is calculated from the Ergun's equation using the mean particle diameter. The minimum fluidization velocity very much depends on the particle size and the particle size distribution, and the smallest particles in a mixture influence significantly on the minimum fluidization velocity (Jayarathne and Halvorsen, 2009). In the simulations with Barracuda, the particle size distribution is included, and the simulations are therefore giving a more realistic value for the minimum fluidization velocity.

Figure 11 shows the distribution of the agglomerates (red) in the bed at minimum fluidization velocity (0.055 m/s) and at velocity 0.085 m/s. The volume fraction of

agglomerates is 26%. At minimum fluidization velocity, the agglomerates are uniformly distributed in the gasifier. However, when the gas velocity is increased to 0.085 m/s, the agglomerates segregates towards the bottom of the bed. Segregation of agglomerates can have unfortunate consequences for the gasification process, as it can lead to defluidized zones with high temperature at the bottom of the bed. This can trigger the formation of more and also larger agglomerates.
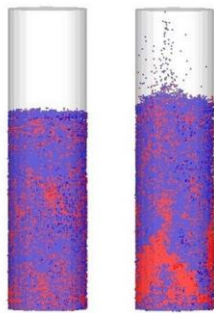
Figure 11. Distribution of agglomerates in the bed at gas velocity 0.055 m/s and 0.085 m/s.

Figure 12 shows the comparison of the flow behavior in the bed without agglomerates (left) and with 26% agglomerates (right) at gas velocity 0.101 m/s. In the sand bed, the bubbles seem to be well distributed in the bed, whereas in the agglomerated bed it seems like the gas (bubbles) is channeling through the bed in the center. The channeling of the air will cause bad mixing and limited mass and heat transfer. It is crucial to run the gasifier at temperatures well below the ash melting temperature to avoid agglomeration of the bed material. It is also important to operate the gasifier well above the minimum fluidization velocity to avoid defluidization and segregation of the larger particles if agglomeration occurs. Reduction in pressure drop over the bed indicates that formation of agglomerates has occurred.

Figure 12. Distribution of particles in bed without agglomerates (left) and with agglomerates (right). Gas velocity 0.101 m/s.
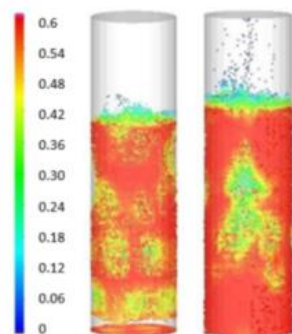
5 Conclusion

The objective of this study was to develop a model to predict the flow behavior in a fluidized bed gasifier. A CPFD model is developed using the commercial software Barracuda VR. Experiments were performed using a cold model of a fluidized bed reactor to study the flow behavior and determine the minimum fluidization velocity. Additional experiments were performed in a lab-scale biomass gasifier 735°C to determine the pressure drop over the bed.

The CPFD model was first validated against experimental data from the cold fluidized bed. Grid resolution tests, tests with different time steps, and tests with different drag models were performed, and the final model gave good agreement with the experimental results. The average deviation between the experimental results and the simulations regarding pressure drop and minimum fluidization velocity was 6% and 10% respectively. The model was further tested against the results from the biomass gasifier, and it gave a good fit for the pressure drop over the bed at gas velocity 0.20 m/s. A few modifications were needed to be able to simulate an agglomerated fluidized bed gasifier. Glicksman's scaling rules were used to scale up the lab-scale gasifier to a full-scale gasifier. Simulations using the modified model were performed to study the flow behavior in a full-scale gasifier with agglomerates. The pressure drop over the bed decreased with increasing mass of agglomerates. The minimum fluidization velocity for the bed with 26 vol % agglomerates was about 10% higher than for the sand bed. The bubble shape and bubble frequency changed with the fraction of agglomerates in the bed. This may be due to gas channeling and segregation of agglomerates. The developed CPFD model is capable of predicting the effect of agglomerates in a fluidized bed gasifier.

References

C. E. Agu, C. Pfeifer, M. S. Eikeland, L. A. Tokheim, and B. M.E. Moldestad. Measurement and characterization of biomass mean residence time in an air-blown bubbling fluidized bed gasification reactor, *Fuel*, 253: 1414-1423, 2019.

W. S. Amarasinghe, C. K. Jayarathna, B. S. Ahangama, B.E.M. Moldestad, and L. A. Tokheim. Experimental study and CFD modelling of fluidization velocity for Geldart A, B and D particles. *International Journal of Modelling and Optimization*, 7(3): 152-156, 2017

M. Balland, K. Froment, G. Ratel, S. Valin, J. Roussely, R. Michel, J. Poirier, Y. Kara, and A. Galnares. Biomass ash Fluidized-Bed Agglomeration: Hydrodynamic Investigations. *Waste Biomass,* 8: 2823-2841, 2017.

J. Bandara, B. M. E. Moldestad, and M. S. Eikeland. Analyzing the Effects of Geometrical and Particle Size Uncertainty in Circulating Fluidized Beds using CPFD

Simulation, *Linköping Electronic Conference Proceedings*, 170: 182-189, 2019.

P. Basu. Biomass gasification, pyrolysis and torrefaction, 2nd ed. Elsevier Inc, 2013.

R Chhabra, L. Agarwal, and N. Sinha. Drag on non-spherical particles: an evaluation of available methods, *Powder Technology*, 101(3): 288-295, 1999.

J. Chladek, C. K. Jayarathna, and B. M. E. Moldestad. L. A. Tokheim, Fluidized bed classification of particles of different sixe and density, *Chemical Engineering Science,* 177: 155–162, 2018.

N. C. I. S Furuvik, R. Jaiswal, and B. M. E. Moldestad. Flow behavior in an agglomerated fluidized bed gasifier, *International Journal of Energy and Environment*, 10(2): 55-64, 2019a.

N. C. I. S Furuvik, R. Jaiswal, R. K. Thapa, and B. M. E. Moldestad. CPFD model for prediction of flow behavior in an agglomerated fluidized bed gasifier", *International Journal of Energy Production and Management*, 4(2): 105-114, 2019b.

N. C. I. S Furuvik, R. Jaiswal, and B. M. E. Moldestad. Experimental study of agglomeration of grass-pellets in fluidized bed gasification, *WIT Transactions on Ecology and Environment*, 246: 9-17, 2020.

L. Glicksman. Scaling relationships for fluidized beds, *Chemical Engineering Science*, 39(9): 1373-1379, 1984.

L. Glicksman, M. Hyre, and P. Farrell. Dynamic similarity in fluidization, *International Journal of Multiphase Flow*, 20: 331-386, 1994.

A. Haider and O. Levenspiel. Drag coefficient and terminal velocityof sperical and nonspherical particles, *Powder Technology*, 58(1): 63-70, 1989.

R. Jaiswal, C. E. Agu, R. K. Thapa, and B. M. E. Moldestad. Study of fluidized bed regimes using Computational Particle Fluid Dynamics. *Linköping Electronic Conference Proceedings*, 153: 271-276, 2018

C. K. Jayarathna and B. M. Halvorsen. Experimental and computational study of particle minimum fluidization velocity and bed expansion in a bubbling fluidized bed. SIMS 50 Conference, Fredericia, Denmark: Technical University of Denmark (ISBN 978-87-89502-88-5)-285290, 2009.

C. K. Jayarathna, B. M. E.Moldestad, and L. A. Tokheim. Validation of results from Barracuda® CFD modelling to predict minimum fluidization velocity and pressure drop of Geldart A particles, *Linköping Electronic Conference Proceedings,* 138: 76-82, 2017.

J. F. Richardson and W.N. Zaki, Sedimentation and fluidization: Part I. *Chemical Engineering Research and Design*, 75: 82–100, 1997.

D. M. Snider. An incompressible three-dimensional multiphase particle-in-cell model for dense particle flows. *Journal of Computational Physics*, 170(2): 523-549, 2001.

R. Turton and O. Levenspiel. A short note on the drag correlation for spheres, *Powder Technology*, 47(1): 83-86, 1986.

C. Wen and Y. Yu. Mechanics of fluidization, *Chemical Engineering Progress Symposium Series*, 62: 100–111, 1966.

M. Öhman and A. Nordin. The Role of Kaolin in Prevention of Bed Agglomeration during Fluidized Bed Combustion of Biomass Fuels. *Energy & Fuels* 14(3): 618-624, 2000.

Simulation of entrained flow gasification reactor with Multi Phase Particle in Cell (MP-PIC) approach

Ramesh Timsina   Rajan K. Thapa   Britt M. E. Moldestad   Marianne S. Eikeland

Department of Process, Energy and Environmental Technology, University of South-Eastern Norway, Porsgrunn,
{ramesh.timsina, rajan.k.thapa, britt.moldestad, marianne.eikeland}@usn.no

Abstract

Entrained flow gasification is a promising technology for conversion of biomass into valuable fuels and chemicals. Residues and byproduct formed during a gasification process possess a significant challenge prior to the production of synthetic fuel (biofuel). Present work focuses on the simulation of an entrained flow gasification reactor in Barracuda, which is based on the Computational Particle Fluid Dynamics (CPFD) modelling. The model is validated against experimental gas compositions reported in the literature. The model was used to study the flow behavior as well as the distribution of product gases and temperature inside the reactor. Simulations showed zones of high and low-temperature regions suggesting different reactions zones such as a partial combustion zone near a fuel injector followed by a gasification zone. The flow behavior inside the reactor shows zones of recirculation, spreading and the fast flowing zone. Results from the product gas distributions inside a reactor supports the reason behind the zones with different temperature.

Keywords: entrained flow, biomass gasification, CPFD, MP-PIC approach

1 Introduction

World economy primarily depends upon the use of fuels, dominated by the use of fossil fuels compared to the use of renewable fuels. Fossil fuels has a share of 81% in the total energy consumption (Dudley, 2018). Due to the negative impact of the use of the fossil fuels, clean and efficient energy sources are getting prioritized in the energies and climate policies across the globe (Solorio and Jörgens, 2020). Biomass energy is a carbon-neutral fuel due to its sustainable life cycle. Besides this, low amounts of N_2 and S in biomass generates low emissions of NOx and SOx. Biomass has been one of the economic and efficient energy sources for the humankind for many years. Biomass energy covers about 14% of the global energy demands and plays an important role in replacing the fossil fuel (Bandara et al., 2018). Biomass gasification is a key technology for the conversion of biomass into syngas, a mixture of CO and H_2. The produced syngas can be used as a source for the production of bio fuels, valuable chemicals as well as for the heat and power generation.

Nearly 25% of the global emission in 2016 were generated by transport sector, out of which air- and road transportation accounts for 86% (Guo, 2020). Fossils fuels primarily drive the transport sector. Among the different alternatives, second generation biofuels via gasification and catalytic conversion is a promising technology. Integration of biofuels from biomass does not require major infrastructure modification, which makes it very relevant to conservative industries such as aviation and marine (Guo, 2020).

Due to the high content of volatile matters, thermal gasification often encounters technical challenges related to tar formation. This problem can be overcome by operating at high temperatures (> 1100°C), which promotes tar-reforming reactions (Llamas et al., 2020). Entrained flow (EF) biomass gasification reactors meet this requirements and typically operate at high temperatures (1300-1500°C) and high pressure (25-30 bar) (Molino et al., 2016).

EF gasification reactors can operate both in a slagging and non-slagging mode. Slagging mode EF biomass gasifiers are more flexible due the ability to melt the ash formed during the operation. Also, the EF gasifiers have high carbon conversion efficiency as compared to the fluidized bed gasifiers (Weiland et al., 2013).

In addition, solid fuel particles have typically very short residence time (2-3 sec) (Qin, 2012). Therefore, smaller particles of typically around a few hundred microns are needed to achieve good heat transfer and mixing (Guo, 2020). Pre-treatment of biomass particles to achieve particle size of a few hundred microns of feedstocks requires a high amounts of energy. Also, the less reactive products, i.e. soot and char, formed during the devolatilization steps limits the complete conversion of fuel. Therefore, it is crucial to model EF reactors accurately to the increase the overall efficiency.

Figure 1 depicts the different processes occurring inside a gasifier. The main chemical reactions for the biomass gasification process are listed below (R1-R7). Heat is supplied to the reactor during the primary pyrolysis/devolatilization, which gives the volatiles and char. Volatiles consists of non-condensable gases such as CO, CO_2, CH_4, H_2 and condensable tar.
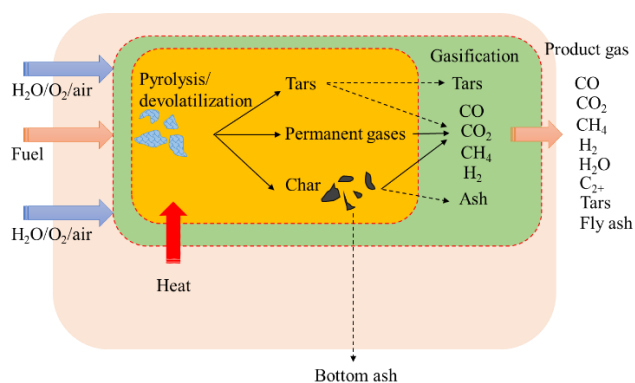
Figure 1. Process occurring in a gasifier: pyrolysis/devolatilization of fuel and gasification/reformation of the resulting gases and char

After initial decomposition, a variety of gas-solid and gas-gas reactions take place. Tars formed during the EF gasification cracks into light hydrocarbons such as CH_4, C_{2+}. Liu et al. have demonstrated that the reaction rate of char gasification is relatively high at a temperature range of 1273-1673K (Liu et al., 2006).

The hydrodynamics and the reactions are quite complex, which limits the optimization of a gasifier performance. It is difficult to study the hydrodynamics and reactions from the experimental tests. Simulation using computational fluid dynamics is becoming an important tool to study these parameters. During this study, a CPFD model is developed for the Pressurized Entrained Flow Biomass Gasification plant (PEBG) operated by Weiland et al. (2013).

1.1 Previous works

Wu et al. (2010) have studied EF coal slurry gasifiers. A three-dimensional numerical model based on the probability density function was developed and the simulation results agree well with the industrial data. Chen et al. (2012) have developed a numerical methods for the prediction of the coal gasification in an EF gasifier. The model particularly focused on the influence of the injection pattern and provide an accurate prediction for syngas formation. Abani and Ghoniem (2013) have developed a model for coal-fed EF gasifiers using large-eddy simulations and Reynolds-averaged Navier Stokes models. The model shows that the unsteady mixing is critical for the estimation of the product gas and carbon conversion. Kumar and Ghoniem (2013) have developed a multiscale model for EF gasifiers to investigate the effects of particle grinding size on carbon conversion. Fine grinding accelerated the char conversion under diffusion-control conditions, whereas there is not noticeable effects under kinetic-control operation.

Due to the complexity of the EF gasification reaction and the limitation of the computational power, the above mentioned model were often simplified to two dimensional or semi three dimensional. Most of the simulations were based on the steady state simulations. There were also limited information about the particle temperatures, carbon content and locations for the discrete particles (Liang et al., 2020).

Liang et al. (2020) have developed a CPFD simulation model for an EF gasification reactor. The detailed particle information and residence time were studied. The rapid expansion from a tracer injector and fast reactions plays an important role in forming the particle distribution zone in the gasifiers.

Thus, further understanding of the reactor hydrodynamics and the transient behavior of the reactor is crucial. This paper will gives information about the transient behavior and the reactor hydrodynamics.

2 Numerical model

There are two distinct approaches in modelling of gas-solid flows in an EF reactor: Eulerian-Eulerian (EE) and Eulerian-Lagrangian (EL) approach. EE modelling defines the gas and solid phases as continuous phases (interpenetrating continua). It lacks the detailed transient information of the two-phase interactions and does not accounts the particle size distribution of the slid phase. The EL approach models the solid phase as discrete elements and the motion of the individual particles is tracked by using Newton's law of motion. The fluid-particle, particle-particle, and particle-wall interaction as well as the particle size distribution is taken into account (Thapa et al., 2014). Thus, EL modelling requires a high computer power to calculate these interactions.

Multi-Phase Particle-In-Cell (MP-PIC) modelling was developed by considering a computational particle as a group of particles (called parcels) with the same size, density, residence time, velocity etc. Parcels are modelled in the discrete frame and the particle interaction are modelled in the Eulerian frame. Fluid particles are solved with an Eulerian set of equations. This reduces the computational costs for discrete modelling of the solid particles.

The main governing equations for CPFD simulations are based on MP-PIC approach and are described by Snider et al. (Snider, 2001; Snider and Banerjee, 2010).

Biomass undergoes devolatilization after the introduction into the reactor. Biomass is then decomposed into char particles and gases at the reactor temperature in the absence of oxygen. Equation 1 defines the global reaction for the devolatilization process (Authier and Lédé, 2013).

$$Wood \xrightarrow{heat} H_2, CO, CO_2, CH_4, chars(s), tars \quad (1)$$
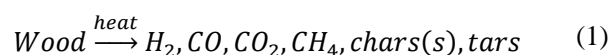
The minor elements such as sulphur and nitrogen are neglected and all the tar formed during the process converts into CO, CO_2 and CH_4. The heavier hydrocarbons such as C_2H_2, C_2H_4 were neglected to

make the model as simple as possible. The devolatilization rate is defined as (Wu et al., 2010):

$$\frac{dm_p}{dt} = -A_v\, T \exp\left(\frac{E_v}{T}\right)(m_p - m_{c,a}) \qquad (2)$$

Where, m_p is the particle mass and $m_{c,a}$ is the mass of char and ash in the particle. The activation energy (E_v) and the pre-exponential factor (A_v) are 3945.15 K^{-1} and 2.1×10^5 s^{-1} respectively (Wu et al., 2010).
Char and ash are the main components of the particles after the devolatilization process. Char reacts with the gases present inside the reactor (heterogeneous reactions). The main reaction involving char gasification are as follows:

$$2C + O_2 \rightarrow 2CO \qquad (R1)$$
$$C + CO_2 \rightarrow 2CO \qquad (R2)$$

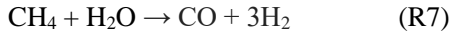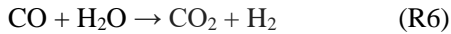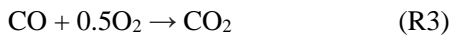The reaction rate for the heterogeneous reactions is determined by the intrinsic reaction rate and the diffusion rate. The expression for the intrinsic reaction rate and the diffusion rate can be found in the study of Wu et al. (2010). Further, char reactivity plays an important role in determining the reaction rate of these equations. The reactivity of char is given by:

$$r_m = -\frac{1}{m_c}\frac{dm_c}{dt} = \frac{1}{(1-x_c)}\frac{dx_c}{dt} \qquad (3)$$

Where m_c and x_c are the mass of carbon contained in the sample and its conversion rate at time t (Gómez-Barea and Leckner, 2010).

A series of homogeneous reaction occurs inside the reactor. Five major global reaction were considered for this study.

$$CO + 0.5O_2 \rightarrow CO_2 \qquad (R3)$$
$$H_2 + 0.5O_2 \rightarrow H_2O \qquad (R4)$$
$$CH_4 + 1.5\,O_2 \rightarrow CO + 2H_2O \qquad (R5)$$
$$CO + H_2O \rightarrow CO_2 + H_2 \qquad (R6)$$
$$CH_4 + H_2O \rightarrow CO + 3H_2 \qquad (R7)$$

The reaction rates for these reactions are listed in Table 1.

Table 1. Reaction rate kinetics (Timsina et al., 2020)

Reactions	Reaction rate (mol.m^{-3}.s^{-1})
R1	$4.34\times10^7 m_s T \exp\left(\frac{-13590}{T}\right)[O_2]$
R2	$1.12\times10^8 m_s P^{0.31}\theta_f \exp\left(\frac{-29518}{T}\right)[CO_4]$
R3	$5.62\times10^{12}\exp\left(\frac{-16000}{T}\right)[CO][O_2]^{0.5}$
R4	$5.69\times10^{11}\exp\left(\frac{-17610}{T}\right)[H_2][O_2]^{0.5}$
R5	$5.0118\times10^{11}\exp\left(\frac{-24357}{T}\right)[CH_4]^{0.7}[O_2]^{0.8}$

R6	$7.68\times10^{10} T \exp\left(\frac{-36640}{T}\right)[CO]^{0.5}[H_2O]$
R7	$3\times10^5\exp\left(\frac{-15042}{T}\right)[CH_4][H_2O]$

The temperature for the heterogeneous reactions were taken as a weighted average with 75% particle temperature and 25% gas temperature.

Barracuda includes the model for both gas-solids and gas-wall heat transfer as well as radiation model. It also has different built-in drag models (Software, 2016).

3 Computational model

The EF reactor simulated in this work is the same reactor constructed and operated by Weiland et al. (2013). The reactor diameter is 0.52m and the height is 1.67m as shown in Figure 2.
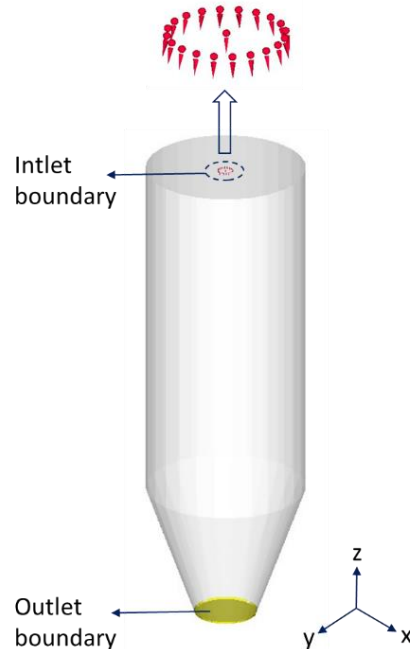
Figure 2. Schematic diagram of the CPFD model and its boundary conditions

A simulation model was developed in the Barracuda VR software. The reactor was modelled as an open cylinder with a conical shaped outlet at the bottom. The Wen-Yu-Ergun drag model was used for this work. A burner was modelled as an injector at the top center of the reactor. Other sets of injectors (20) were place concentrically outside of the fuel entrance burner. Oxygen required for gasification was supplied through these injector boundary parameters. The operating conditions of the reactor are shown in Table 2. Table 3 gives the properties of the biomass used during their study.

A total of 87300 real cells were generated using the inbuilt mesh generator available in Barracuda. In the CPFD simulations, the number of computational particles is controlled by a parameter called the number

density (Software, 2016). Number density was set to 125000 to achieve a smoother and healthier particle feed for the system. This gives the particle to cell ratio of about 10:1.

Table 2. Experimental test conditions performed by Weiland et al. (2013)

Particle size, µm	100
Fuel feeding rate, kg/h	40
Total N_2 inlet, kg/h	14.4
O_2 inlet, kg/h	26.6
O_2 equivalence ratio	0.44
System pressure, bar	1.94

Injection boundary conditions were used to define the inflow of fuel and gasifying agents along with nitrogen into the reactor. Accuracy of the injection boundary is not affected by the mesh sizes of the geometry. The angle of expansion of the injection boundary was set to 20° but it is significantly dependent upon the gas behavior inside the reactor. A pressure boundary was defined at the bottom of the reactor to allow the outflow of the gas and the solid particles.

Table 3. Properties of the soft stem wood used by Weiland et al. (2013)

Proximate analysis (wt. %, dry)	
Fixed carbon	15.1
Volatile matter	84.5
Ash	0.4
Ultimate analysis (wt. %, dry)	
C	50.90
H	6.30
O	42.4
N	0.10
S	0.006
Cl	0.02

4 Results

The developed model was simulated for 50 seconds. The average gas composition were taken as the time average over final 20 seconds of simulations. The obtained results were compared with the results from an experiment performed by Weiland et al. (2013)[1].

The average molar composition of the produced gas on nitrogen free dry basis is 0.457 of CO, 0.275 of H_2, 0.226 of CO_2 and 0.038 of CH_4. Table 4 shows that the simulation results agree well with the experimental results. The mole percentage of CH_4 in the experiment also includes the mole percentage of C_2H_2 (0.3) and C_2H_4 (0.1).

Table 4. Comparison between the simulation and experimental results (mole percentage on nitrogen free dry basis)

	Product gas species			
	CH_4	CO	CO_2	H_2
Simulation	3.8	45.7	22.6	27.5
Experiment	2.7	48.5	21.1	27.8

As the EF gasification reactors operate at a high temperature, it was desired to monitor the reactor temperature. The gas temperature distribution inside the reactor is presented in Figure 3.
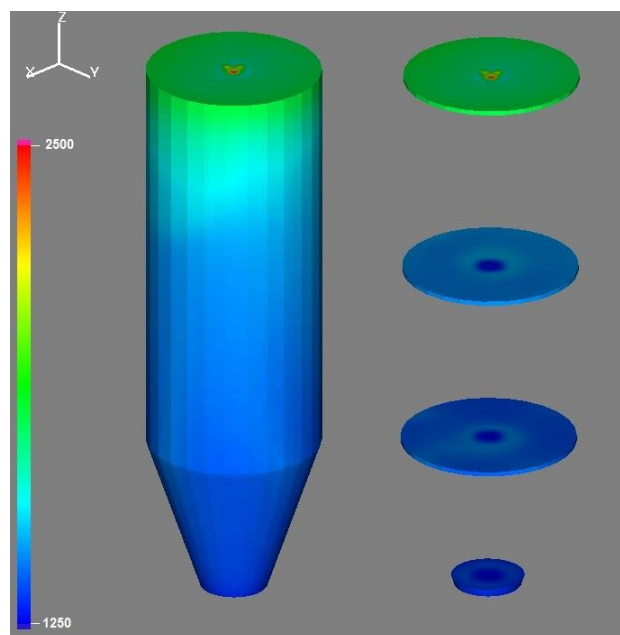
Figure 3. Gas temperature (K) distribution inside the reactor

It can be seen from Figure 3 that the temperature around the fuel injector is comparatively higher than in the rest of the reactor. Often the reactor injector up to the burners are purged with nitrogen to avoid the burning of biomass before the burner (Weiland et al., 2013). The reactor temperature at different cross sections (right) shows that the temperature distribution becomes uniform with an increase in the reactor depth.

The product gas composition was monitored along the height of the reactor. Figure 4 shows the mole fractions of CO, H_2 and CO_2 along the height of the reactor. There is a high concentration of CO and H_2 along the center of the reactor. From the distribution of the CO, it can be seen that there are dead spots at the top corner of the reactor. This gives rise to the uneven distribution of the gas components and the temperature inside the reactor. The gas distribution is similar along the radial direction except for in the top region (high temperature region as can be seen from Figure 3).

---

[1] Results are taken from the experiment on 14 February.
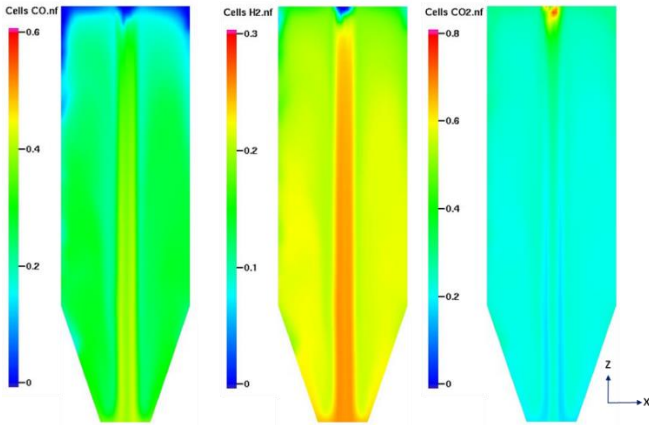
Figure 4. Gas composition along the reactor (mole fraction at t = 30 sec)

The high temperature close to the fuel burner (fuel injector) gives rise to high concentration of CO_2 and low concentration of CO and H_2. As the mixture of biomass and the pyrolysis gas move down, several chemical and physical transformation of the biomass occurs resulting in the product gas composition as shown in the Figure 4.

Therefore, it is important to know the fluid velocity and direction inside the reactor. Figure 5 shows the instantaneous fluid velocity distribution. It can be seen from the figure that recirculation of gas occurs near the wall of the reactor. The gas velocity in the central region gradually increases as it flows downward in the gasifier. The rapid gas expansion as well as recirculation is due to the expansion effects of the injection nozzles. Due to the jet velocity along the axial direction, expansion in radial direction is high compared to the axial direction. This is in agreement with the published result by Liang et al. (2020), where the reactor has three distinct flowing zones, i.e. the recirculation zone, the spreading zone and the fast flowing zone. The flow direction is random except in the middle of the reactor. This behavior has a great influence on the particle flow as well as the overall conversion efficiency of the process.

A summary of the results from the experiment can be obtained from the published article by Weiland et al. (2013). The article also compares the results from the different gasification technologies. A gasification process with higher concentration of CH_4 in the product gas is more suitable for power generation as well as for Substitute Natural Gas (SNG) production.
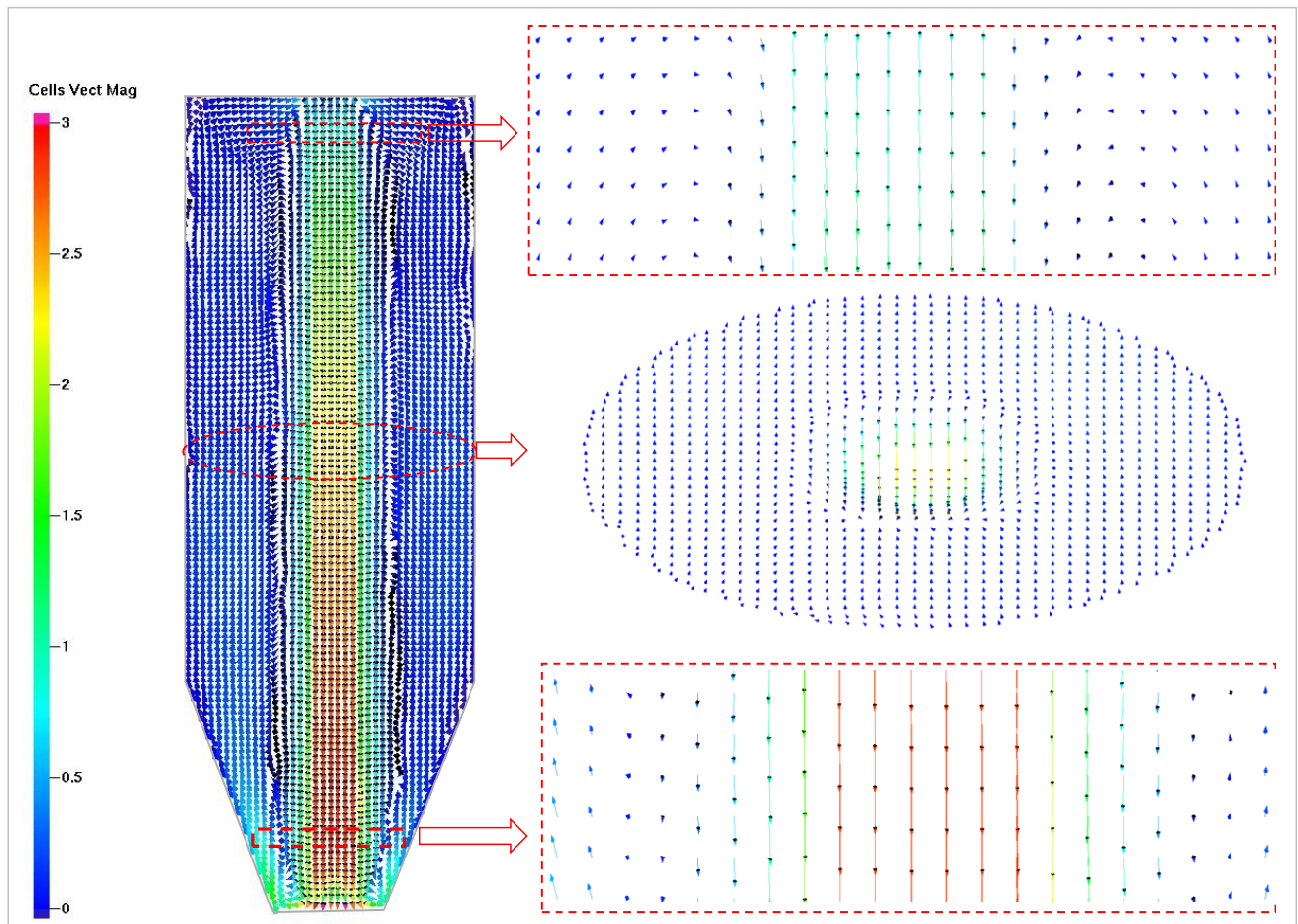
Figure 5. Gas speed distribution at t = 30 sec.

The H_2/CO ratio is an important parameter for the conversion of syngas into synthetic fuels. The ratio of H_2/CO will vary depending upon the synthesis route. For example, the low temperature Fischer-Tropsch synthesis (FT Synthesis) requires H_2/CO ratio in the range of 1.7 - 2.15 depending upon the catalyst, while the ratio is approximately 1.05 for FT synthesis at higher temperature (Weiland et al., 2013). Therefore, syngas requires shifting towards high H_2 content prior to the fuel synthesis irrespective of the gasification technology.

Higher operating temperature for the EF reactor reduces the amounts of tar and heavier hydrocarbons in the product gas compared to the other gasification technologies. This potentially reduces the cost for the extensive syngas cleaning prior to fuel synthesis. However, the gasification pressure needs to be high enough to make the conversion process economically feasible. The convective and radiative losses from the reactor also plays an important role for the thermal efficiency of the plants. Dry biomass powder was gasified during the experiments in PEBG gasifier.

However, Brown et al. (1986) have shown that the premixing of the coal with steam or coal with moisture gave high concentration of H_2, but lower CO/CO_2 ratio decreases the carbon conversion. This could be due to the reduced gasification temperature.

Therefore, it is important to characterize (pros and cons) different alternatives before selecting a suitable conversion technology for the conversion of the biomass into biofuels via gasification. The difficulty of understanding the hydrodynamics as well as the reaction chemistry during an experiment can be studied by developing a simulation model. A CPFD model can give a detail insight of the reaction operating conditions, which in turn help a lot for the optimization and design of the EF reactor.

5 Conclusion

A CPFD simulation model was developed in Barracuda using the MP-PIC modelling approach. The model was used to simulate a pressurized entrained flow biomass gasification reactor operated by Weiland et al. (2013). The composition of the product gases obtained from the model agree well with the experimental results. The average molar composition of the produced gas on nitrogen free dry basis is 0.457 of CO, 0.275 of H_2, 0.226 of CO_2 and 0.038 of CH_4. An accurate prediction of the reactor performance is a challenging task, which is investigated in this study. A simple CFD model is presented in this work, which needs testing in different conditions and the authors believe that the model will be of use in the development and design of the entrained flow biomass reactor.

The gas expansion played a significant role for the particle speed and direction inside a reactor. Certain groups of particles in the center of the reactor has higher velocity and lower residence time. Other groups of the particles are recirculated giving a different flow direction and velocity. The CO_2 concentration is highest and the CO and H_2 concentration is lowest at the fuel injector.

Selection of suitable technology for the production of syngas prior to the synthetic fuel production depends upon different criteria such as biomass feed, desired syngas quality, capacity and costs. Entrained flow reactors are best suited for a feed with small particles at large capacity, at high temperatures and high pressures. Entrained flow reactors give cleaner syngas compared to fluidized and fixed bed reactors, which potentially reduces the cost for the extensive syngas cleaning prior to fuel synthesis.

Acknowledgements

Abbreviations

CPFD	Computational Particle Fluid Dynamics
EF	Entrained flow
EE	Eulerian-Eulerian
EL	Eulerian-Lagrangian
FT Synthesis	Fischer-Tropsch synthesis
MP-PIC	Multi-Phase Particle-In-Cell
PEBG	Pressurized Entrained Flow Biomass Gasification plant
SNG	Substitute Natural Gas

References

N. Abani and A. F. Ghoniem. Large eddy simulations of coal gasification in an entrained flow gasifier. *Fuel* 104: 664-680, 2013. doi.org/10.1016/j.fuel.2012.06.006.

O. Authier and J. Lédé. The image furnace for studying thermal reactions involving solids. Application to wood pyrolysis and gasification, and vapours catalytic cracking. *Fuel* 107: 555-569, 2013. doi.org/10.1016/j.fuel.2013.01.041.

J. C. Bandara, B. ME. Moldestad and M. S. Eikeland. Analysing the effect of temperature for steam fluidized-bed gasification of biomass with MP-PIC simulation. *International Journal of Energy and Environment* 9(6): 529-542, 2018.

Blaine W. Brown, L. Douglas Smoot and Paul O. Hedman. Effect of coal type on entrained gasification. *Fuel* 65(5): 673-678. 1986. doi.org/10.1016/0016-2361(86)90363-7.

C-J. Chen, H. Chen-I., and W-H. Chen. Numerical investigation on performance of coal gasification under various injection patterns in an entrained flow gasifier. *Applied energy* 100: 218-228. 2012. doi.org/10.1016/j.apenergy.2012.05.013.

B. Dudley. BP statistical review of world energy. *BP Statistical Review*, London, UK, accessed June 30 2020: 2018.

A. Gómez-Barea and B. Leckner. Modeling of biomass gasification in fluidized bed. *Progress in Energy and Combustion Science* 36(4): 444-509. 2010. doi.org/10.1016/j.pecs.2009.12.002.

N. Guo. Modelling of reacting multi-phase flow for biomass gasification. *Doctoral theses at NTNU* 1-224. 2020.

M. Kumar and A. F. Ghoniem. Application of a validated gasification model to determine the impact of coal particle grinding size on carbon conversion. *Fuel* 108: 565-577. 2013. doi.org/10.1016/j.fuel.2013.02.009.

Y. Liang, C.Y. Guo, X. Zhao, Q. Qin, Y. Cheng, and L. He. CPFD simulation on particle behaviour in an entrained-flow gasifier. *Clean Energy* 4(1): 75-84. 2020. doi.org/10.1093/ce/zkz032.

H. Liu, C. Luo, S. Kato, S. Uemiya, M. Kaneko andT. Kojima. Kinetics of CO_2/Char gasification at elevated temperatures: Part I: Experimental results. *Fuel Processing Technology* 87(9): 775-781. 2006. doi.org/10.1016/j.fuproc.2006.02.006.

Á. D. G. Llamas, N. Guo, T. Li, R. Gebart, T. Løvås, and K. Umeki. Morphology and volume fraction of biomass particles in a jet flow during devolatilization. *Fuel* 278: 118241. 2020. doi.org/10.1016/j.fuel.2020.118241.

A. Molino, S. Chianese, and D. Musmarra. Biomass gasification technology: The state of the art overview. *Journal of Energy Chemistry* 25(1): 10-25. 2016. doi.org/10.1016/j.jechem.2015.11.005.

K. Qin, A. D. Jensen, P. A. Jensen and W. Lin. Entrained Flow Gasification of Biomass. *Technical University of Denmark, Department of Chemical Engineering*. 1-224. 2012.

D. M. Snider. An incompressible three-dimensional multiphase particle-in-cell model for dense particle flows. *Journal of computational physics* 170(2): 523-549. 2001. doi.org/10.1006/jcph.2001.6747.

D. M. Snider and S. Banerjee. Heterogeneous gas chemistry in the CPFD Eulerian–Lagrangian numerical scheme (ozone decomposition). *Powder Technology* 199(1): 100-106. 2010. doi.org/10.1016/j.powtec.2009.04.023.

C. Software. Barracuda Virtual Reactor User Manual. *CPFD Software LLC*. Albuquerque. 2016

I. Solorio and H. Jörgens. Contested energy transition? Europeanization and authority turns in EU renewable energy policy. *Journal of European Integration* 42(1): 77-93. 2020. doi.org/10.1080/07036337.2019.1708342.

R. Thapa, C. Pfeifer and B. M. Halvorsen. Modeling of reaction kinetics in bubbling fluidized bed biomass gasification reactor. *Int. J. Energy Environment* 5: 35-44, 2014.

R. Timsina, R. K. Thapa, B. M. E. Moldestad and M. Eikeland. Experiments and computational particle fluid dynamics simulations of biomass gasification in an air-blown fluidized bed gasifier. *Int. J. of Energy Prod. & Mgmt.* 5(2): 102-114. 2020. doi.org/10.2495/EQ-V5-N2-102-114.

F. Weiland, H. Hedman, M. Marklund, H. Wiinikka, O. Öhrman, and R. Gebart. Pressurized oxygen blown entrained-flow gasification of wood powder. *Energy & Fuels* 27(2): 932-941. 2013. doi.org/10.1021/ef301803s.

Y. Wu, J. Zhang, P. J. Smith, H. Zhang, C. Reid, J. Lv, and G. Yue. Three-dimensional simulation for an entrained flow coal slurry gasifier. *Energy & Fuels* 24(2): 1156-1163. 2010. doi.org/10.1021/ef901085b.

A Data-Driven Approach for the Prediction of Subcooled Boiling Heat Transfer

Jerol Soibam[1]   Ioanna Aslanidou[1]   Konstantinos Kyprianidis[1]   Rebei Bel Fdhila[1,2]

[1]Mälardalen University, Västerås, Sweden, `{jerol.soibam, ioanna.aslanidou, konstantinos.kyprianidis}@mdh.se`
[2]Hitachi ABB Power Grids, Västerås, Sweden, `rebei.bel_fdhila@hitachi-powergrids.com`

Abstract

In subcooled flow boiling, heat transfer mechanism involves phase change between liquid phase to the vapour phase. During this phase change, a large amount of energy is transferred, and it is one of the most effective heat transfer methods. Subcooled boiling heat transfer is an attractive trend for industrial applications such as cooling electronic components, supercomputers, nuclear industry, etc. Due to its wide variety of applications for thermal management, there is an increasing demand for a faster and more accurate way of modelling.

In this work, a supervised deep neural network has been implemented to study the boiling heat transfer in subcooled flow boiling heat transfer. The proposed method considers the near local flow behaviour to predict wall temperature and void fraction of a sub-cooled mini-channel. The input of the network consists of pressure gradients, momentum convection, energy convection, turbulent viscosity, liquid and gas velocities, and surface information. The output of the model is based on the quantities of interest in a boiling system i.e. wall temperature and void fraction. The network is trained from the results obtained from numerical simulations, and the model is used to reproduce the quantities of interest for interpolation and extrapolation datasets. To create an agile and robust deep neural network model, state-of-the-art methods have been implemented in the network to avoid the overfitting issue of the model. The results obtained from the deep neural network model shows a good agreement with the numerical data, the model has a maximum relative error of 0.5 % while predicting the temperature field, and for void fraction, it has approximately 5 % relative error in interpolation data and a maximum 10 % relative error for the extrapolation datasets.

Keywords: deep neural network (DNN), CFD, machine learning (ML), sub-cooled boiling, heat transfer.

1 Introduction

Subcooled flow boiling exists when the bulk liquid temperature remains below its saturation value, but the surface is hot enough for bubbles to form. Subcooled boiling is generally observed in micro-channel heat sinks, due to high heat transfer rates and maintaining relatively low wall temperature under high subcooled conditions. Due to the nature of high heat transfer mechanism in subcooled boiling, it plays a paramount role in enhancing the thermal efficiency of a system in industrial applications. However, the physical process that occurs during the subcooled boiling process is a complex phenomenon, and it remains a major challenge to predict the heat transfer behavior correctly in the boiling regime. One of the first attempts was made by Mohammed (1977), where they used multiple experimental data of different fluids to derive a correlation for low and high subcooled regions to predict the heat transfer coefficients. Kandlikar (1998) carried out an extensive review of the subcooled boiling heat transfer correlation, and he also introduced the region of void flow, where the convective heat transfer plays an important role due to void fraction. Yin et al.(2000) examined subcooled boiling heat transfer of R-134a in a horizontal annular duct, where they showed that increased subcooling had a huge effect in the bubble diameter. Through this work, they proposed an empirical correlation for boiling heat transfer coefficient and bubble departure diameter.

More recently, Multiphase Computational Fluid Dynamics (MCFD) models have been used for modeling boiling heat transfer and regarded as promising tools to understand the underlying physic in boiling flows. Jakobsen et al.(2005) reported that among the available modeling techniques, the Eulerian multi-fluid approach is most commonly used for bubble simulation. The Eulerian multi-fluid approach treats each phase as interpenetrating continua and relies on an ensemble averaging of the multiphase Navier-Stokes equations (Ishii, 1975). To solve the two-fluid model, a closure equation has to be introduced in the conservation equations for the individual phase, and the accuracy of the model highly depends on the closure equations (Besagni et al., 2018). These closure terms are combination of mechanistic models and empirical correlation, including nucleation site density, bubble departure diameter, bubble departure frequency, and heat flux. A comprehensive review on

MCFD from Cheung et al.(2014) stated that there is no single combination of empirical correlations that can provide satisfactory predictions for all the quantities of interest for a variety of input conditions.

Machine learning (ML) algorithms provide an agile framework for specific problems, let it be medical data or fluid mechanics data. These algorithms can be used as an additional tools to optimize, understand the correlation that exist in data, speed up simulation and, better understand the underlying physics that occurs in fluid flow simulation or heat transfer problems. In recent years ML techniques are rapidly making an inroad towards fluid mechanics and heat transfer domain, and this is becoming possible due to high-performance computing and the advancement of experimental techniques. Ansari et al.(2020) successfully implemented Artificial Neural Network (ANN) for multiphase flow to predict the dynamic features including pressure, velocity, and the evolution of phase fraction. They claimed that their model was capable of predicting within few minutes under 10 % error. Alic et al.(2019) tested various ML algorithms (Genetic algorithms, ABC algorithms, Support vector machine (SVM), Decision Tree (DT), and Multi-Layer Perceptron (MLP)) to predict the heat flux of pool boiling heat transfer. According to their study SVM regression model gave the best prediction. Kumar et al.(2018) use ANN as an inverse model to estimate unknown heat flux in fin heat transfer with a maximum error of 0.41 %. They used surrogate data generated from Asymptotic CFD as an input to the ANN. Hobold and da Silva(2019) applied supervised and unsupervised ML techniques to study boiling heat transfer mechanisms, concerning bubble dynamics and bubble morphology on a wire. They demonstrated that the ANN-based model could quantify heat transfer using only direct and indirect visual information of the boiling phenomenon. Moreover, they used these models to infer heat flux in real-time using Raspberry Pi. Buist et al.(2019) used ML algorithms to learn the closure term for stratified multiphase flow in channels, from an unsteady high-fidelity simulation data. Then, the ML model was used to establish a functional relation between the two-fluid model and the closure terms, which served as the source term to the two-fluid model. Ma et al.(2015) used a neural network to fit Direct Numerical Simulation (DNS) data to develop closure relations for the average two-fluid equations. Once the network was trained it was used for different initial velocity and void fraction. They observed that the average equations with the neural network closure model were capable of reproducing the main aspect of the DNS results. The aforementioned, methods focus on addressing the heat transfer behavior in pool boiling, multiphase flow in a channel, or deriving a closure model with the aid of ML techniques. However, according to the authors best knowledge, there is no study related to sub-cooled boiling heat transfer in mini-channel using Deep learning techniques to predict the void fraction and the temperature field.

This work focuses on the investigation of a Deep Neural Network (DNN) model to estimate the quantities of interest for subcooled boiling in a mini-channel. Data for training this model is obtained from CFD simulations and the data are then divided into training, validation, and testing datasets. The model used in this work is tailored to predict the quantities of interest for varying heat flux and inlet velocities. This model has 16 input features, and 2 output features, namely wall temperature and void fraction. Overall, this study shows that the DNN model can capture the physics and non-linear behaviors that exist in subcooled boiling heat transfer in mini-channel. Once the DNN model is trained and validated it is used to predict the QoIs for a new case for interpolation and extrapolation datasets, and it predicted with exceptional accuracy. From the predictive nature of the trained DNN model and its predictive speed, it opens up the possibility to use such a model for the design phase in thermal management for subcooled boiling systems.

2 Numerical Method for data generation

The data used for training the Deep NN in this work are obtained from applying Eulerian two-fluid method to model the subcooled boiling flow in a mini channel. Numerical simulations are performed using open-source software OpenFOAM. The three conservation equation of incompressible Navier-Stokes equations are solved to generate the data. The mass conservation equation for each phase can be written as following:

$$\frac{\partial \alpha_k \rho_k}{\partial t} + \nabla . (\alpha_k \rho_k \mathbf{U}_k) = \Gamma_{ki} - \Gamma_{ik} \tag{1}$$

where k is the phases, α is the void fraction, ρ is the phase density, $\mathbf{U}$ is the phase velocity and Γ is the mass transfer rate per unit volume. The momentum conservation equation for each phase:

$$\frac{\partial \alpha_k \rho_k \mathbf{U}_k}{\partial t} + \nabla . (\alpha_k \rho_k \mathbf{U}_k \mathbf{U}_k) = -\alpha_k \nabla p + \mathbf{R}_k + \mathbf{M}_k +$$
$$\alpha_k \rho_k \mathbf{g} + (\Gamma_{ki} \mathbf{U}_i - \Gamma_{ki} \mathbf{U}_k) \tag{2}$$

where ∇p is the pressure gradient, $\mathbf{R}$ is the combined turbulent and laminar stress term, calculated based on the Reynolds analogy, $\mathbf{g}$ is the gravitational acceleration, and $\mathbf{M}$ is the interfacial momentum transfer which accounts for the drag forces. The energy transport equation is written in terms of specific energy h for each phase k as following:

$$\frac{\partial \alpha_k \rho_k h_k}{\partial t} + \nabla . (\alpha_k \rho_k \mathbf{U}_k h_k) = \alpha_k \frac{Dp}{Dt} + \nabla \left(\alpha_k D_{t,k}^{eff} \nabla h_k \right)$$
$$+ \Gamma_{ki} h_i - \Gamma_{ik} h_k + Q_{wall,k} \tag{3}$$

The standard $k-\varepsilon$ turbulence model is used to account for turbulence dispersed flow in the vapor phase. To account for the turbulence flow behavior of the continuous phase, the Lahey $k-\varepsilon$ (Lahey Jr, 2005) turbulence model is used for the liquid phase. The $k-\varepsilon$ turbulence model is adopted since it is valid for fully turbulent flows and computationally inexpensive also due its robust nature for pipe flows. The boiling model used here is based on the well known RPI model (Kurul, 1991), where the total heat flux $q_w^{''}$ on the wall is divided in three components, evaporation heat flux $q_{w,e}^{''}$, forced convection heat flux $q_{w,c}^{''}$ and quenching heat flux $q_{w,q}^{''}$. The total applied heat flux can be written as follows:

$$q_w^{''} = q_{w,e}^{''} + q_{w,c}^{''} + q_{w,q}^{''} \qquad (4)$$

For this simulation active nucleation site density is define on the surface where total heat flux is applied as per the correlation (Benjamin and Balakrishnan, 1997):

$$N_a = 218 \mathrm{Pr}_l^{1.63} \Delta T_{sup}^3 \gamma^{-1} \theta^{-0.4} \qquad (5)$$

where Pr_l is the liquid Prandtl number, ΔT_{sup} is the wall superheat, γ is a coefficient taking into account the liquid and heated surface thermophysical properties and θ is a coefficient taking into account the heated surface roughness and the system pressure.

The bubble departure diameter is calculated based on the semi-empirical model of Ünal (Ünal, 1976), given as following:

$$d_{dep} = \frac{2.42 \ 10^{-5} p^{0.709} a}{\sqrt{b\phi}} \qquad (6)$$

where a and b are the model coefficients, taking into account the working fluid and the heated surface thermophysical properties, and ϕ is a parameter controlled by the local flow velocity.

The closure term for the bubble departure frequency is calculated according to the mechanistic model of Brooks and Hibiki (Brooks and Hibiki, 2015) as:

$$f_{dep} = \frac{C_{fd} \mathrm{Ja}_w^{0.82} N_T^{-1.46} \rho^{*-0.93} \mathrm{Pr}_{sat}^{2.36}}{d_{dep}^2} \qquad (7)$$

where C_{fd} is the model coefficient depending on the size of the channel where boiling occurs, N_T is a dimensionless temperature, ρ^* is a dimensionless density ratio, Ja_w is a modified Jacob number and Pr_{sat} is the liquid Prandtl number evaluated at the corresponding saturation temperature. With the initial and boundary conditions specified along with closure terms, the solver is able to predict the boiling heat transfer accurately and data are generated.

2.1 Data Extraction

The 2D computational domain of the mini-channel used for simulation is shown in the Figure 1. The channel has a total height of 0.6 m along the y-axis and 0.003 m along the x-axis. To avoid the influences of the boundary, non-developed flow, and to account, only the surface where
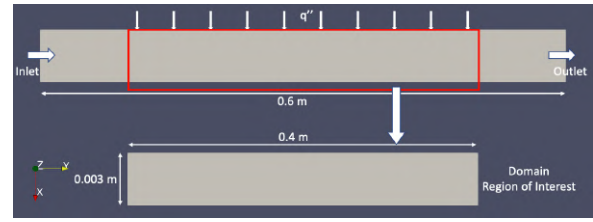
Figure 1. Domain region of interest

Table 1. Input features used for training the network.

Input Features	Feature Expressions
Pressure gradient	$\frac{\partial \langle p \rangle}{\partial x}$ $\frac{\partial \langle p \rangle}{\partial y}$
Momentum convection	$\frac{\partial \langle p \rangle \langle u \rangle \langle u \rangle}{\partial x}$ $\frac{\partial \langle p \rangle \langle u \rangle \langle v \rangle}{\partial x}$ $\frac{\partial \langle p \rangle \langle u \rangle \langle v \rangle}{\partial y}$ $\frac{\partial \langle p \rangle \langle v \rangle \langle v \rangle}{\partial y}$
Energy convection	$\frac{\partial \langle p \rangle \langle T \rangle \langle u \rangle}{\partial x}$ $\frac{\partial \langle p \rangle \langle T \rangle \langle v \rangle}{\partial y}$
Total heat flux	q_{total}
Velocity inlet	U_{inlet}
Pressure inlet	p_{inlet}
Temperature inlet	T_{inlet}
Ambient pressure	p_{amb}
Fluid and gas viscosity	$\mu_l \ \mu_g$
Non-dimensional x and y axis	x^*, y^*

Table 2. Output features (quantities of interest).

Output Features	Feature Expressions
Wall Temperature	T_{wall}
Void Fraction	α

heat flux is applied, domain region of interest (ROI) is selected for data extraction. The number of cells in the ROI is 321x26 resulting in 8,346 data points for each case. The domain axis (x and y) is further converted into a non-dimensional number so that the model is not constrained to learn based on the height of the channel and applicable for other channel lengths. In total 106 simulations have been performed for inlet velocity ranging from 0.05 ms^{-1} to 0.2 ms^{-1} and heat flux ranging from 1000 Wm^{-2} to 40000 Wm^{-2}. 100 cases are used for training and validation, further, the data are split into 80 % percent for training and 20% for validation purpose. The remaining 6 cases are used for evaluating the model performance.

The selected feature inputs obtained from CFD simulations are shown in the Table1. The inputs are chosen based on their influence on the quantities of interest. The

quantities of interest chosen in this work are void fraction which controls the pressure drop and the flow regimes, and wall temperature which is relevant for cooling. The output of the deep neural network model is shown in the Table 2. Before feeding the data into the network, the input features and output features are normalized between 0 and 1. This way the ML algorithms can learn better since the scale of data is very sparse. It is also worth mentioning that the data are randomly shuffled with a random instance of 8000 data points for every reshuffle. Shuffling the data ensures low variance and avoids the risk of creating batches that are not representative of the overall dataset while training the model.

3 Deep Neural Network Architecture

DNN is a type of Artificial Neural Network, where the number of hidden layers is more than one layer. The first and last layer serve as input and output layers similar to that of an ANN. The DNN architecture used in this work is shown in Figure 2. The network is trained using the backpropagation technique, meaning there are two phases involved in training the network. The first phase is called the forward phase. Here the signals from the input layer (training data) are propagated along the hidden layers with series of non-linear transformations controlled by weights and biases, followed by a nonlinear activation function until the output layer is reached.

$$h_1 = g(W_1^T x + b_1)$$
$$..$$
$$h_5 = g(W_5^T h_4 + b_5) \tag{8}$$
$$\hat{y} = g(W_6^T h_5 + b_6)$$

$$g(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases} \tag{9}$$

The activation function $g(x)$ used in this work is rectified linear units (ReLU), where h stands for hidden layers, W is the weights of each layers, and b is biases of the network.

A loss function is defined while training the network to measure the error between the predicted value $\hat{y}$ and the target value y. For this work, the mean squared error (MSE) is used as the loss function to compute the error then the error gradient is used to compute the new weights and biases of the neurons. Based on the computed gradients, the weights and biases are updated using Adaptive Moment Estimation (Adam) (Kingma and Ba, 2014) optimization technique in a backward direction with a learning rate of Lr = $1e^{-4}$. Hence, the second phase is called the backward phase.

The following measures have been taken to avoid overfitting while training the network. First and foremost L_2 regularization term is introduced in the loss function. Secondly, callback functions are defined to save only the best
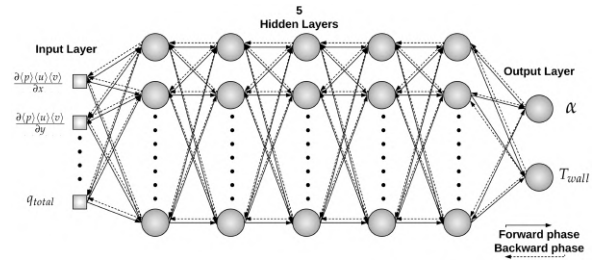
Figure 2. Architecture of the 5 hidden layer DNN.

weights of the network, early stopping of the training if the validation loss does not improve in the next epoch. Epoch is a term used when training a ML model, and it indicates the number of passes of the entire training dataset the ML algorithm has completed.

$$L_{loss} = MSE(y, \hat{y}) + \lambda \sum_{i=1}^{N} w_i^2 \tag{10}$$
$$L_2 norm$$

Where λ is a positive hyperparameter that influences the regularization term, N is the total number of data. While training the network other metrics such as, RMSE and coefficient of determination (R^2) along with loss MSE function were defined to measure the accuracy of the model. Figure 3 shows the training RMSE and validation RMSE of the model trained for 500 epochs.
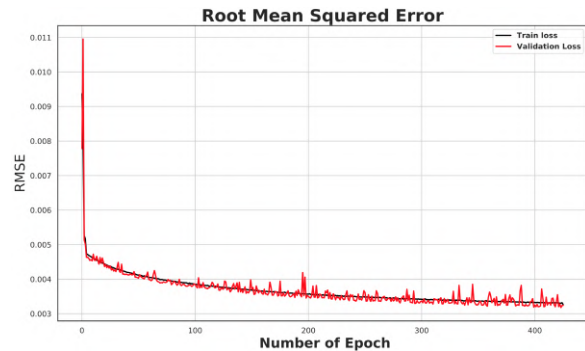
Figure 3. Training and validation RMSE error for 500 epochs.

In this work, the open source deep learning library Tensorflow 1.14 from google is used for constructing the model. The model is trained using GPU which allowed faster training compared to CPU.

4 Results and Discussion

4.1 Validation dataset

Before validation or evaluation of the model, the best weights of the model saved by the callback functions during training are first loaded. This confirms that the model weights used for validation and evaluation are the best one, and it is not an overfitted model. The errors in temperature, and void fraction values predicted using the DNN

Figure 4. Predicted error on validation dataset for DNN model.

Table 3. Statistics table of the validation dataset. Val: Validation, VF: Void Fraction, Temp: Temperature

Dataset	R^2		RMSE	
	VF	Temp	VF	Temp
Val	0.998	0.995	0.0059	0.134

model from that of CFD values is illustrated in Figure 4. It is seen that the error lies around zero value for both the fields. Looking at the number of counts it can be interpreted that the DNN model slightly under predicts for the void fraction, since there are more counts on the negative side.

To get more insight of the model and its robustness, regression chart with relative error % is shown in Figure 5. The relative error here is defined as: $(Predicted_{value} - CFD_{value}) * 100/CFD_{value}$. The solid 45° black line is the true line where DNN predicted values would perfectly match the CFD results. The blue dashed line stands for relative percentage error, the relative error is different for temperature and void fraction due to differences in scale. It can be further observed from the figure that the relative error for predicted temperature is under 0.5 % error, which clearly shows that the DNN model can capture the temperature variation well. Whereas, in the case of the validated void fraction, the relative error is about 5 %, and it is seen that there are few outliers near-zero void fraction. The outliers present in the plots are statistically insignificant when compared to the number of data points (83460) present in the plot. The quantitative values of the DNN model on the
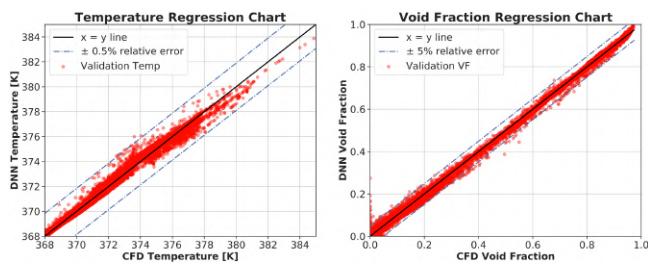
Figure 5. Comparison of CFD value and DNN value for the validation dataset.

validation dataset is presented in Table 3. From the table, it can be depicted that the model fits very well with the validation data and has a very low value of RMSE.

4.2 Test dataset

4.2.1 Interpolation datasets

Interpolation dataset here means that the chosen Q = 15000 Wm^{-2} at velocity inlet u =0.05 ms^{-1} lies between the range of training data, but this data were not used for training nor validation. Once the model has been

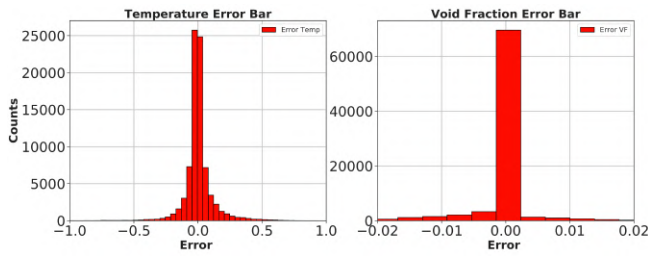validated, the unseen test data are used to evaluate the model. Figure 6 shows the temperature field obtained from a CFD result, the temperature field predicted using the DNN model, and the relative percentage error between them. It is noticeable from the figure that the error ranges between ± 0.3 %, which implies the DNN model can capture the complex nature, and the model could be used for predicting unseen cases. However, the model slightly over predicts the temperature field near the wall of the channel, and under predict around the center of the channel. To understand furthermore and to closely observe the performance of the model in near-wall region Figure 7 is plotted, which shows the trend of void fraction and wall temperature along the channel. From the figure it can be further noted that the DNN model is capable of predicting the sub-cooled boiling regimes closely to that of CFD values.
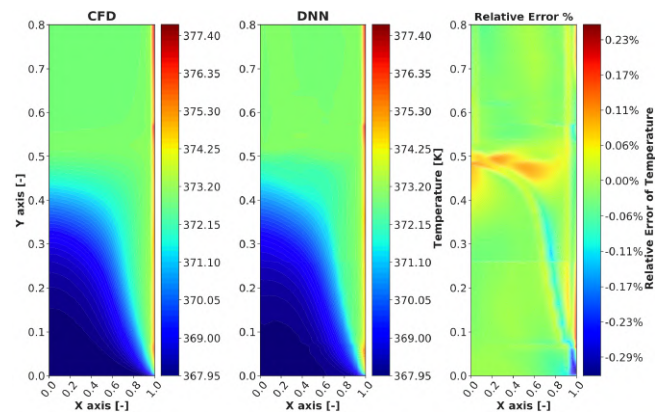
Figure 6. Temperature field for CFD, DNN and relative error % for Q =15,000 Wm^{-2} at u = 0.05 ms^{-1}.
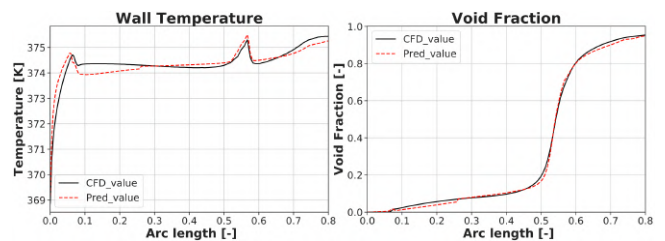
Figure 7. CFD, DNN wall void fraction and wall temperature for Q =15,000 Wm^{-2} at u = 0.05 ms^{-1}. Arc length: The non-dimensional length along the y-axis near the wall of the minichannel.

The void fraction field in the channel obtained from CFD and predicted results from DNN is presented in Fig-
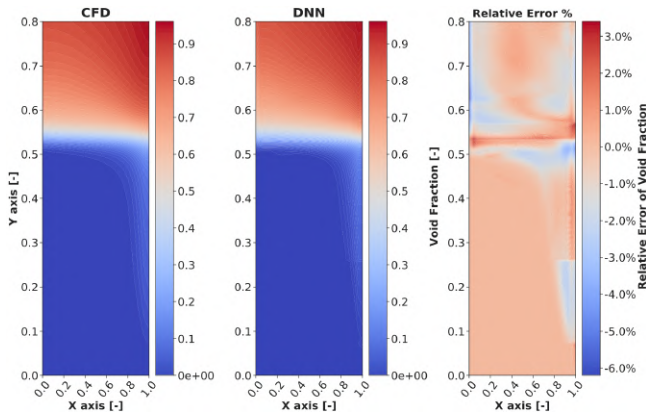
Figure 8. Void fraction field for CFD, DNN and relative error % for Q =15,000 Wm^{-2} at u = 0.05 ms^{-1} .

Figure 10. Temperature field for CFD, DNN and relative error % for Q =40,000 Wm^{-2} at u = 0.2 ms^{-1} .

ure 8. Here it can be seen that the relative error ranges from 3.2 % to -6 %, and this error is mostly concentrated near the non-dimensional arc length of 0.5 where the void fraction suddenly starts to increase. This sudden change in void fraction can be observed in Figure 7, and it can be observed that temperature is closely related to this phenomenon. One explanation could be due to the presence of a hotspot, caused by the formation of a bigger bubble or a large number of bubbles near the wall resulting in a sudden increase in void fraction and temperature.

Figure 9 shows how well the model fits on the interpolation test data. From the plot, it can be spotted that the trend of error i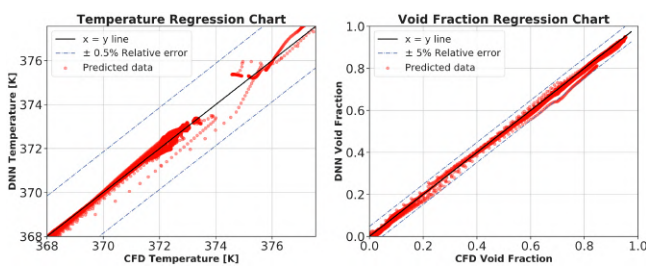s similar to that of the validation dataset, but the number of outliers is lower because the test data is 10 times less than that of validation data. Again from the temperature regression chart, it is noticeable that there is an offset in predicted temperature around 374 K, but the relative error is still under 0.3%. With the above-presented results and from Table4, it is possible to conclude that the DNN model can predict well for interpolation cases.

ranges from Q= 1000 to 29,000 Wm^{-2} and u = 0.05 to 0.2 ms^{-1} , while the data used for extrapolation have a heat flux value of Q =40,000 Wm^{-2} at u =0.2 ms^{-1} . The motive was to see if the DNN model has learned sufficiently the underlying physics of subcooled boiling from the data provided during training, and hence reproduce the physics on unseen cases away from the training region.

The temperature field obtained from CFD data and the predicted value by the DNN model is shown in Figure 10. From the figure it can be observed that the DNN model follows the same pattern of temperature distribution along the channel. The next plot shows how far is the predicted value from the CFD value, and it is worth noticing that the relative error is still under $\pm$ 0.5 % error. The possible explanation for this behavior could be that the DNN model has the capability of learning from the local features and intrinsic pattern present in the training data, and it can extrapolate the data points to explain the physical process of subcooled boiling temperature.

Figure 9. CFD, DNN regression chart of void fraction and temperature for Q =15,000 Wm^{-2} at u = 0.05 ms^{-1} .

4.2.2 Extrapolation dataset

The dataset used for extrapolation is an extreme case, meaning this data is far away from any data that is used during training of the model. It is to be noted that the extrapolation data sets are already available through CFD calculation but it is not used for training the DNN. The maximum heat flux and velocity used during training

Figure 11. CFD, DNN wall void fraction and wall temperature for Q =40,000 Wm^{-2} at u = 0.2 ms^{-1} .

The void fraction and the wall temperature near the wall region are illustrated in Figure 11. From the plot, it can be noted that the DNN model over predicts until the arc length of 0.24 and it starts to under predicts until 0.75 arc length then follows the pattern of CFD data. From this, it is possible to conclude that the DNN model finds it difficult to capture the void fraction when its value is equal to zero. The temperature profile near the wall, however, follows a good trend compared to CFD data.

Table 4. Performance of the DNN model on test dataset, VF: Void Fraction, Temp: Temperature, Inter: Interpolation, Extra: Extrapolation. *: Results presented in the paper.

Cases	U ms^{-1}	Q Wm^{-2}	RMSEP		R^2		Error(%)	
			VF	Temp	VF	Temp	VF	Temp
Inter*	0.05	15 000	0.0088	0.13	0.999	0.995	3.64 −6.09	0.31 −0.26
Inter	0.075	17 500	0.0066	0.129	0.999	0.996	4.3 −4.5	0.26 −0.36
Extra	0.1	30 000	0.011	0.23	0.998	0.98	9.48 −8.33	0.617 −0.303
Inter	0.15	19 000	0.0025	0.08	0.96	0.99	2.4 −1.4	0.121 −0.27
Extra	0.15	30 000	0.0084	0.196	0.98	0.99	7.07 −9.29	0.4 −0.414
Extra*	0.2	40 000	0.013	0.278	0.982	0.987	10.8 −10.5	0.58 −0.47
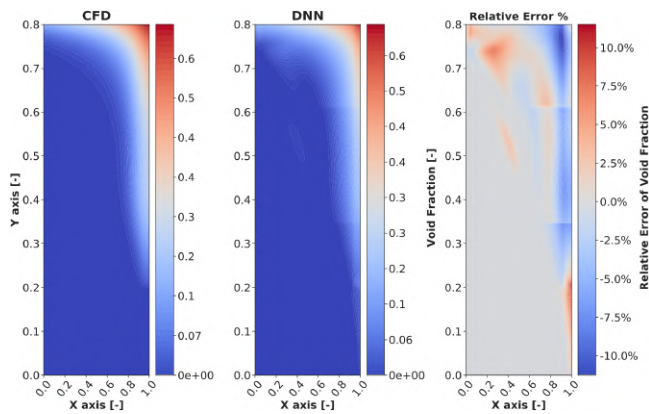
Figure 12. Void fraction field for CFD, DNN and relative error % Q =40,000 Wm^{-2} at u = 0.2 ms^{-1} .

points around 376 K, which accounts for 0.5 % relative error in the model.
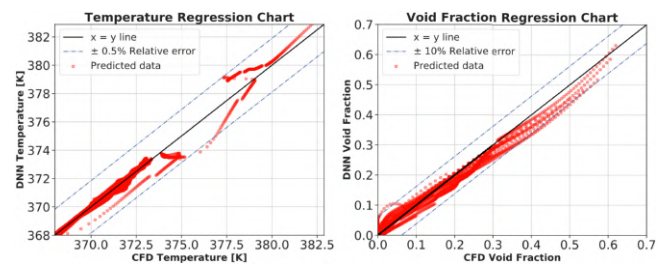
Figure 13. CFD, DNN regression chart of void fraction and temperature for Q =40,000 Wm^{-2} at u = 0.2 ms^{-1} .

The CFD domain and the domain predicted using the DNN model is indicated in Figure 12. From the relative error plot it is noticeable that the maximum relative error is in the range of ± 10.5 %. These errors are mostly concentrated near the wall, and the upper region of the channel. The possible explanation for such a behavior may be due to the lack of data for such high heat flux value in the training data and the physical nature of the problem in these region is difficult to predict resulting in poor prediction. Nevertheless, the DNN model still follows the trend of CFD void fraction value, and it can predict the whole domain under 10 % relative error. Figure 13 shows how well the model fits the extrapolation data, for void fraction around 90 % of the data lies between ± 5% relative error to that of the true line (solid black line) and remaining 10 % of the data points are the sparse one which inflicts the maximum error in the model. Looking at temperature data distribution, there are some non-linear data

The model performance for all the test data is presented in Table 4. The test data consist of interpolation and extrapolation dataset. It is noticeable that the model is evaluated, on quite a variety of datasets. This provides a strong indication of how generalized the model used in this work is. It is worth noticing that the RMSEP values are well below 0.3, and the coefficient of determination score is about 0.98 on average. This shows how well the model fits the data. From the presented results, it is possible to conclude that the model performs very well when predicting the temperature field in the domain with a maximum relative error of 0.6 % for both interpolation and extrapolation datasets. However, the models still lacks good accuracy when it comes to predicting the void fraction field, especially near the wall for the extrapolation datasets. Nonetheless, the model can still follow the trend of CFD values and predict well for the interpolation datasets, and for extrapolation datasets that are not far away from the values used for training.

5   Conclusion and Future Work

In this work, a supervised deep neural network model based on the backpropagation technique is implemented to study the subcooled boiling heat transfer. The training data are obtained from CFD simulations from varying heat fluxes and inlet velocities. In total 106 cases were simulated, out of which 80 cases were used for training the network, 20 cases for validation purposes, and the remaining 6 cases for evaluating the model. After training, the performance of the DNN model was validated using the validation dataset, and then the model performance was evaluated using test datasets. The model showed good level of accuracy while predicting temperature and void fraction field for both interpolation and extrapolation dataset. In the case of extreme extrapolation, the model performance deteriorated while predicting the void fraction field with a maximum error of 10.8%. Nonetheless, the model showed an exceptional level of accuracy while predicting the temperature field with a maximum error of 0.6%. From the results obtained in this work, it is demonstrated that the DNN model is capable of capturing the subcooled boiling pattern, and it is applicable for interpolation and extrapolation data extraction of the quantities of interest. Moreover, the DNN model trained on the local flow features has acceptable generalization property, and it can be used as a promising tool to help improve the predictive speed of MCFD solvers. Therefore, the future work will focus on improving the predictive nature of the DNN model for void fraction field, and for quantifying the uncertainty present in the DNN model.

Acknowledgements

References

Erdem Alic, Mehmet Das, and Onder Kaska. Heat flux estimation at pool boiling processes with computational intelligence methods. *Processes*, 7(5):293, 2019.

A Ansari, SSH Boosari, and SD Mohaghegh. Successful implementation of artificial intelligence and machine learning in multiphase flow: Two case studies of cfd problems. *J Pet Environ Biotechnol*, 11:401, 2020.

RJ Benjamin and AR Balakrishnan. Nucleation site density in pool boiling of saturated pure liquids: effect of surface microroughness and surface and liquid physical properties. *Experimental Thermal and Fluid Science*, 15(1):32–42, 1997.

Giorgio Besagni, Fabio Inzoli, and Thomas Ziegenhein. Two-phase bubble columns: A comprehensive review. *ChemEngineering*, 2(2):13, 2018.

Caleb S Brooks and Takashi Hibiki. Wall nucleation modeling in subcooled boiling flow. *International Journal of Heat and Mass Transfer*, 86:183–196, 2015.

JFH Buist, Benjamin Sanderse, Yous van Halder, Barry Koren, and GertJan van Heijst. *Machine Learning for Closure Models in Multiphase-Flow Applications.* PhD thesis, Master's thesis, Eindhoven University of Technology, 2019.

SCP Cheung, S Vahaji, GH Yeoh, and JY Tu. Modeling subcooled flow boiling in vertical channels at low pressures–part 1: Assessment of empirical correlations. *International Journal of Heat and Mass Transfer*, 75:736–753, 2014.

Gustavo M Hobold and Alexandre K da Silva. Visualization-based nucleate boiling heat flux quantification using machine learning. *International Journal of Heat and Mass Transfer*, 134:511–520, 2019.

Mamoru Ishii. Thermo-fluid dynamic theory of two-phase flow. *STIA*, 75:29657, 1975.

Hugo A Jakobsen, Håvard Lindborg, and Carlos A Dorao. Modeling of bubble column reactors: progress and limitations. *Industrial & engineering chemistry research*, 44(14):5107–5151, 2005.

Satish G Kandlikar. Heat transfer characteristics in partial boiling, fully developed boiling, and significant void flow regions of subcooled flow boiling. 1998.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Harsha Kumar and Gnanasekaran Nagarajan. A synergistic combination of asymptotic computational fluid dynamics and ANN for the estimation of unknown heat flux from fin heat transfer. *Alexandria Engineering Journal*, 57(2):555–564, 2018.

N Kurul. On the modeling of multidimensional effects in boiling channels. *ANS. Proc. National Heat Transfer Con. Minneapolis, Minnesota, USA, 1991*, 1991.

Richard T Lahey Jr. The simulation of multidimensional multiphase flows. *Nuclear Engineering and Design*, 235(10-12):1043–1060, 2005.

Ming Ma, Jiacai Lu, and Gretar Tryggvason. Using statistical learning to close two-fluid multiphase flow equations for a simple bubbly system. *Physics of Fluids*, 27(9):092101, 2015.

Shah Mohammed and Mohammed Shah. A general correlation for heat transfer during subcooled boiling in pipes and annuli. 1977.

HC Ünal. Maximum bubble diameter, maximum bubble-growth time and bubble-growth rate during the subcooled nucleate flow boiling of water up to 17.7 mn/m2. *International Journal of Heat and Mass Transfer*, 19(6):643–649, 1976.

Chih-Ping Yin, Yi-Yie Yan, Tsing-Fa Lin, and Bing-Chwen Yang. Subcooled flow boiling heat transfer of r-134a and bubble characteristics in a horizontal annular duct. *International journal of heat and mass transfer*, 43(11):1885–1896, 2000.

Parameter estimation using Bayesian approach for a model of Pressure Peaking Phenomena-ignited H$_2$ releases.

Agnieszka Weronika Lach   Torbjørn Grande Østby

Process, Energy and Environmental Technology, University of South-Eastern Norway, Kjølnes Ring 53 3916 Norway,
`Agnieszka.Lach@usn.no`
`Torbjorn.G.Ostby@usn.no`

Abstract

The aim of this paper is to estimate the parameters of the model of Pressure Peaking Phenomena (PPP). This project focuses on the investigation of the overpressures arising from the ignited hydrogen releases in 14.9m^3 enclosure (explosive chamber) through a 4mm nozzle. The various ventilation areas and mass flow rates were applied in 31 tests. The controlled variables for experiments are mass flow rate (MFR, g/s), ventilation area (A$_v$, m^2) and time of hydrogen releases (t, s). The Bayesian approach was implemented in the parameter estimation using Markov chain Monte Carlo method for simulations. The discharge coefficient and heat loss coefficient has been analyzed and gave by posterior distribution.

Keywords: pressure peaking phenomena, hydrogen safety, Bayesian analysis, parameter estimation, MCMC.

1   Introduction

Hydrogen releases in confined spaces are bringing new hazards into underground transportation systems. Unexpected hydrogen releases in confined spaces can result in dynamic overpressures with characteristic peaks. The phenomenon called pressure peak phenomena is distinct for hydrogen and occurs while introducing gas with lower density than the gas inside the enclosure (Brennan and Molkov, 2018). The characteristic transient pressure can be observed during releases with a high molar flow in combination with a relatively low ventilation area.

In a study by Makarov et al. (2018) a model for hydrogen jet fires from the TPRD (Thermal activated Pressure Relief Devices) was presented, which was demonstrated to be consistent with the experimental observations on in ~1m^3 enclosure. The model predicted much higher overpressures compared to the unignited releases. The model consists of the volumetric mass balance of the gasses in the enclosure during a combustion process. They applied a perfect mix assumption and included adiabatic temperature in the calculations. The results of their simulation for garage-like enclosures showed the risk of enclosure collapse in a few seconds. The parameters used for the simulation were typical TPRD diameter and low ventilation area

(commonly used in UK and France). Further numerical study of pressure peaking from ignited hydrogen releases was performed by Hussein et al. (2018). Their simulations were performed for small scale enclosures (experimental result from 1m^3 enclosure) and real scale (garage-like) for the common use of TPRD diameters. The model used ANSYS ICEM CFD (Computational Fluid Dynamics) to generate geometries for both studied cases. The RANS (Reynolds-Average Navier-Stokes) conservation equations were used for solving energy, mass, species and momentum, turbulent model, radiation, and combustion model. The EDC (Eddy Dissipation Concept) model was used to solve the combustion process and DO (Discrete Oridantes) model to include the radiation process. Results demonstrated the relation between vent area and release rate. The big impact of heat transfer mechanisms in the prediction of the pressure peaking phenomenon was acknowledged. It was concluded that the current TPRD diameters may result in a significantly dangerous situation in under-ventilated enclosures. The heat transfer mechanisms (radiation and conduction) has been investigated (Hussein et al., 2018). The assumption of adiabatic walls showed high over predictions compared to simulations where both radiation and conduction were included. The heat transfer has been computationally demonstrated but due to the small scale of experiments (2 s experimental run) couldn't be physically observed. The heat transfer in the system is found to be important and needs further investigation.

In this study, the model of ignited hydrogen releases developed at USN (University of South-Easter Norway) is presented for the first time. It is based on the problem approach presented by Brennan et al. (2010). The model was validated against large scale experiments in the explosion chamber, allowing to observe high overpressures (over 45 kPa). In this study, we use the Bayesian parameter estimation approach to estimate the model parameters, the discharge coefficient and the heat loss. The Bayesian method combines the information on a discrepancy between the model and the measurements given a set of parameter values (described by the likelihood) and the information available in the literature (described by the prior) through use of the Bayes rule to determine the probable values of the unknown parameters. An important distinction between

the Bayesian method and other classical methods is that it can explicitly and consistently incorporates all the existing knowledge about the unknown parameters. The results of each parameter consistent with the measurements are represented by a posterior distribution. We use a Markov Chain Monte Carlo (MCMC) method to generate samples from the posterior, which can be used for estimation of the unknown parameters. In recent years, the popularity of the MCMC methods has been on the rise (Vrugt, 2016). The prior distribution is chosen based on information found in the literature in order to impose bounds on the parameter space. The prior and posterior densities have been graphically presented as a normal distribution function as in studies on combustion kinetics models (Wang et al., 2020).

The main goal is to create a probability density function (pdf) for parameters giving the most accurate model.

Using Bayesian analysis allows us to understand the parameters' influence on pressure dynamics in the system hence improve the model. The pressure peaking model is designated to be an engineering tool for safety engineers. The accurate parameter estimation plays a significant role in its development.

The investigation of occurring overpressures from ignited hydrogen releases in confined spaces is part of pre-normative research for the safety of hydrogen fuel vehicles and transport through tunnels and similar confined spaces (Hy-Tunnel CS).

2 Case set up and methodology

2.1 Experimental setup and materials

All experiments were performed in a steel explosion with a total volume of 14.9 m^3. The explosion chamber's walls in total have five vents of 80 mm diameter each (0.005027m^2). The vent in the middle of the front wall was used for the hydrogen and propane pipes. Three of the vents were used to vary the passive ventilation area. The flanges were used to fully close/open the vents with gasket ensuring no leakage. The specifics of the releases were obtained with hydrogen mass flow through a stainless pipe outlet located in the center of the chamber's floor to vertically discharge hydrogen jets fires.

The Coriolis mass flow meter and the pneumatic valve were mounted to measure and control hydrogen releases. Oscilloscope Sigma has recorded pressure and

mass flow rate from Coriolis mass flow. The complete overpressure development was measured with Kulite pressure transducer XTM - 190-50A. Oscilloscope Gen3i recorded the overpressure constantly with parallel measurement initiated by the voltage signal (with 25 ks/s). In the table below are listed the uncertainties of measurement.

Table 2. Standard deviations of instruments

Pressure sensor	±1% FSO BFSL 250 point (0.025 sec) filter
Mass flow sensor	±0.5% of a flow rate
Thermocouples type K	±0.75%

2.1.1 Data set

The experiments were designed for three different ventilation areas and with a variety of mass flow rates The purpose of those experiments was to validate the model of PPP. To observe overpressure the mass flow rate of discharged hydrogen into an enclosure has to be relatively high while the ventilation area has to be relatively small (Makarov et al., 2018). The chosen combination of vent area and mass flow rates is based on previous unignited experiments (Lach, 2019) and studies on PPP (Hussein et al., 2018). The pressure peak phenomenon for hydrogen jet fires in the 14.9m^3 enclosure was successfully observed during all 31 experiments. Experiment 11 described in Table 1 will be used to present methods of the Bayesian approach for parameter estimation.

The measured pressure of experiment 11 is presented in Figure 1. The uncertainty of the Kulite sensor together with the sampling time, result in uncertainty -/+ 0.22 kPa.
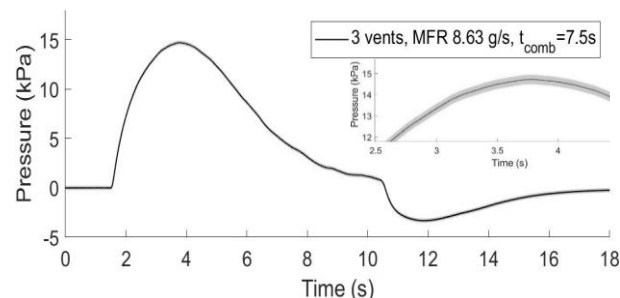
Figure 1. Exp 11; Overpressure for MFR 8.63 g/s with 3 open vent with total vent area 0.0164 m^2. Measurement uncertainty 0.22kPa.

Table 1 Experimental results: H$_2$ releases through 4mm nozzle in 14.9m^3 enclosure with ventilation area: 1 open vent=0.0055m^2, 2 vent open= 0,0109m^2, 3 vents open= 0.0164m^2.

Setup					Measured	
Exp nr	T$_0$ in enclosure [K]	H$_2$ release time [s]	Open vent	Mass flow rate [g/s]	Experimental overpressure (max) [kPa]	Experimental underpressure (max) [kPa]
11	280	7.50	3	8.63	14.70	-3.30

2.2 Model of Pressure Peaking Phenomena-governing equation

The model of pressure dynamics in the enclosure during ignited hydrogen releases (jet fires) is presented. In the calculations, the stoichiometric hydrogen combustion was applied where one mole of hydrogen requires 0.5 moles of oxygen: $H_2+0.5(O_2+3.76N_2)\leftrightarrow H_2O+1.88N_2$. The system of conservation equations described in this chapter provides a solution for the PPP model explaining mass balance and temperature based on the internal energy in the enclosure.

The mass balance for each species in the enclosure is described with the system of equations (1)-(6). The number of moles in the enclosure n_{tot} is a sum of the number of moles of each species in the enclosure $n_{tot}=\Sigma n_i$:

$$\frac{dn_{H_2}}{dt} = \dot{n}_{inH_2} - \dot{n}_{outH_2} + \dot{n}_{rxH_2} \tag{1}$$

$$\frac{dn_{O_2}}{dt} = \dot{n}_{inO_2} - \dot{n}_{outO_2} + \dot{n}_{rxO_2} \tag{2}$$

$$\frac{dn_{N_2}}{dt} = \dot{n}_{inN_2} - \dot{n}_{outN_2} + \dot{n}_{rxN_2} \tag{3}$$

$$\frac{dn_{H_2O}}{dt} = \dot{n}_{inH_2O} - \dot{n}_{outH_2O} + \dot{n}_{rxH_2O} \tag{4}$$

Where $\dot{n}_{in,H_2} = \dot{m}_{in}/M_{H_2}$ (hydrogen mass flow $\dot{m}_{in}$ was measured with coriolis mass flow meter), given the initial conditions (at t=0) $n_{H_2.0} = 0. n_{O_2.0} = 0.21 \cdot n_{tot.0}$. $n_{N_2.0} = 0.79 \cdot n_{tot.0} . n_{H_2O.0} = 0$ and $n_{tot.0} = p_0 \cdot \frac{V}{RT_0}$. If the pressure inside the enclosure will be lower than ambient pressure $\Delta p<0.001$ that means the hydrogen mass flow into the enclosure was closed and $\dot{n}_{in,H_2} = 0$. Then the air from outside the chamber starts to flow into the enclosure through the vent area, means the $\dot{n}_{out,i} = 0$ and:

$$n_{in,} = \frac{C_{in} \cdot A}{M_{air}} \sqrt{2 \cdot \Delta p \cdot \rho_{air}} \tag{5}$$

Where M_{air} is the molecular mass of air and ρ_{air} is the density of air. The C_{in} and A is the discharge coefficient and the vent area, respectively. Using equation (5) the $\dot{n}_{in,i}$ can be calculated: $\dot{n}_{in,O_2} = 0.21 \cdot \dot{n}_{in}$ and $\dot{n}_{in,N_2} = 0.79 \cdot \dot{n}_{in}$ given the assumption that $\dot{n}_{in,H_2O} = 0$.

By balancing the combustion reaction equation, to burn all hydrogen $\dot{n}_{rx,H_2} = -\dot{n}_{in,H_2}$ the oxygen needed will be: $\dot{n}_{rx,O_2} = 0.5 \cdot \dot{n}_{rx,H_2}$. The produced water vapor then will be $\dot{n}_{rx,H_2O} = -\dot{n}_{in,H_2}$, and $\dot{n}_{rx,N_2} = 0$.

Equation (6) based on the steady-state in a compressible energy equation (Bernoulli equation) gives the molar flow out $n_{out,tot}$, which contains all the species in the enclosure from the initial stage and combustion products (no condensation of the water on the walls inside the chamber):

$$n_{out,tot} = C \cdot A \sqrt{\frac{2 \cdot \Delta p \cdot n_{tot}}{V \cdot M_{en}}} \tag{6}$$

Where $n_{tot} = \sum_i n_i$ is the number of moles in the enclosure, the molecular mass in the enclosure is then $M_{en} = \sum_i X_i M_i$ where $X_i = n_i/n_{tot}$ and $\dot{n}_{out,i} = X_i \cdot n_{out,tot}$ can be calculated. The C is the discharge coefficient and A is the ventilation area.

To calculate the pressure in the enclosure (12) the temperature has to be solved first. The internal energy U for the system is equal to:

$$\frac{dU}{dt} = \dot{H}_{in} - \dot{H}_{out} + \dot{Q}_{rx} - \dot{Q}_{loss} \tag{7}$$

And

$$\frac{dU}{dt} = \frac{dn\hat{U}}{dt} \tag{8}$$

With the assumption of specific internal energy $\hat{U}$:

$$\hat{U}_i = \hat{U}_{ref,i} + \int_{T_{ref}}^{T} c_{v,i}dt \tag{9}$$

Where $\hat{U}_{ref,i} = 0$ at $T_{ref} = 298.15K$ and $c_{v,i}$ is the molar heat capacity in a constant volume. Changing the internal energy to the temperature will result in a governing equation of temperature in the enclosure T_{en}:

$$\frac{dT_{en}}{dt} = \frac{1}{\sum_i n_i c_{v,i}} \left(\dot{H}_{in} - \dot{H}_{out} + \dot{Q}_{rx} - \dot{Q}_{loss} - (T - T_{ref}) \cdot \sum_i c_{v,i} \frac{dn_i}{dt} \right) \tag{10}$$

The $\dot{H}_{in} = \sum_i \dot{n}_{in.i} \hat{H}_i$ is the sum of enthalpies of each species at $T = T_{in}$. Using the same analogy the enthalpy of the system at $T = T_{en}$ is the sum of enthalpies: $\dot{H}_{out} = \sum_i n_{out.i} \cdot \hat{H}_i$.

The $c_{v,i}$ was calculated by reducing the heat capacity at constant pressure by the universal gas constant $c_{v,i} = c_{p,i} - R$. The $c_{p,i} \dot{H}_{in,i}$ and $\dot{H}_{out,i}$ where calculated with NASA polynomials (Mcbride et al., 1993) which includes thermodynamic data coefficients and enthalpy of formation $\hat{H}_{f,i}$.

The $\dot{Q}_{rx}$ is the heat of formation. Since the water vapor is the only product $\dot{Q}_{rx} = \dot{Q}_{rx,H_2O} = \dot{n}_{rxH_2O} \cdot (-\Delta \hat{H}_{f_{H_2O}})$ where the $\hat{H}_{f_{H_2O}}$ is the enthalpy of formation of water vapor.

The $\dot{Q}_{loss}$ is the heat loss calculated with the major assumption of a simple heat transfer with no condensation of water: $\dot{Q}_{loss} = h_{loss} \cdot A_{wall} \cdot (T_{en} - T_{wall})$ where A_{wall} is the surface area inside the enclosure and h_{loss} is the heat transfer coefficient. When the pressure inside the enclosure will be lower than ambient pressure $\Delta p < 0.001$ the cold air is entering into enclosure and heat transfer coefficient is assumed to be $h_{loss,2} = 0.5 * h_{loss}$.

The change of the number of moles in enclosure causes temperature change, expressed in equation (10) with $((T - T_{ref}) \cdot \sum_i c_{v,i} \frac{dn_i}{dt})$. The temperature of the wall T_{wall}, was calculated with the major assumption that the whole wall is one thermal mass with the same temperature inside (i.e. no temperature gradient in the wall):

$$\frac{dT_{wall}}{dt} = \frac{\dot{Q}_{loss}}{m_{wall} \cdot C_{steel}} \tag{11}$$

Where m_{wall} is the mass of walls and C_{steel} is the heat capacity of the wall.

Solving the PPP model with conservation equations (1), (2), (3), (4), (10), (11) then allows to calculating pressure inside the enclosure with the ideal gas law (12):

$$p_{en} = \frac{n_{tot} \cdot R \cdot T_{en}}{V} \qquad (12)$$

2.3 Bayesian framework

A Bayesian approach with Monte Carlo simulation was used in estimating the parameters: the discharge coefficient, C, and the heat transfer coefficient, h_{loss}. The purpose of using Bayesian inference is to obtain new recognition about the parameters θ. Whereas standard optimization techniques would estimate the optimal value for these parameters, Bayesian simulation instead estimates a distribution of probable values (based on including observed data and prior knowledge). Using a Bayesian approach makes sense given the uncertainties inherent in experiments and measurements.

The relationship between $\theta = [C, h_{loss}]$ and the measurement data d is given according to Bayes' theorem, equation (13). C and h_{loss} are assumed to be independent.

$$p_{post}(\theta|d, M, I) = \frac{\mathcal{L}(d|\theta, M, I) p_{prior}(\theta|M, I)}{\int \mathcal{L}(d|\theta, M, I) p_{prior}(\theta|M, I) d\theta} \quad (13)$$

$$\propto \mathcal{L}(d|\theta, M, I) p_{prior}(\theta|M, I)$$

Here $p_{post}(\theta|d, M, I)$ is the posterior probability for parameters θ given the measurement data d, M representing the previously described model, and I represents other unspecified information given in this paper, such as the experimental setup. The $p_{post}(\theta|d, M, I)$ expresses the updated knowledge of the θ given the d, M, I. Further, $p_{prior}(\theta|M, I)$ represents the prior probability distribution for parameters θ given the model M and the information I-the initial belief about the θ. Preliminary testing suggested that the value of h_{loss} was between 20 and 40, while the value of C was between 0.6 and 1. The prior distributions used are further detailed in the next subchapter. Finally, $\mathcal{L}(d|\theta, M, I)$ represents the likelihood (probability) of seeing the measurement data d given the parameters θ, the model M, and other information I. The Bayesian calculation will confront the simulation results of the earlier described model as a function of θ and d, M, I to gather the pdf (probability density function) of the θ for the observed data.

Assuming the measurement uncertainties are normally distributed, the likelihood is given by the Gaussian function (Daly et al., 1995). As experiments consist of multiple measurements, this gives us equation (14), where d_i (experimental overpressure) and m_i (simulated overpressures) represent measurement data

and model data at index i (step), respectively. The sigma, σ is the standard deviation of the measurement uncertainty (constant σ, Table 2).

$$\mathcal{L}(d|\theta, M, I) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(d_i - m_i)^2}{2\sigma^2}\right) \quad (14)$$

Having the priors and likelihood allows the posterior pdf to be estimated through Monte Carlo simulation. This was done using the standard Metropolis Markov Chain Monte Carlo algorithm, which is further detailed in algorithm 1.

Algorithm 1

The Metropolis algorithm (adapted from (Kruschke, 2015))

1. **Given:** data d; prior distribution $p_{prior}(\theta|M, I)$; likelihood function $\mathcal{L}(d|\theta, M, I)$; step standard deviation σ; number of steps T
2. Initialize θ^0
3. **For** $i = 1$ **to** T:
4. Sample $\theta' \sim \mathcal{N}(\theta^{i-1}, \sigma_2)$
5. Sample $u \sim \mathcal{U}(0,1)$
6. $p_{move} = \min(1, \frac{\mathcal{L}(d|\theta', M, I) p_{prior}(\theta'|M, I)}{\mathcal{L}(d|\theta^{i-1}, M, I) p_{prior}(\theta^{i-1}|M, I)})$
7. **If** $p_{move} \geq u$: $\theta^i = \theta'$; **else**: $\theta^i = \theta^{i-1}$

This is a sampling algorithm where samples producing a higher probability than the current sample is always accepted, while samples producing a lower probability than the current sample is sometimes accepted, depending on a randomly sampled value. This process is then repeated for a limited number of steps.

However, with high numbers of measurements, the likelihood has a tendency to vanish, as many values between 0 and 1 are multiplied. Using the natural logarithm is a natural way of overcoming this. Using the natural logarithm changes steps 6 and 7 of algorithm 1 in the following way. The fraction in step 6 is changed as shown in equation (15). Here ℓ refers to the loglikelihood, and $\wp_{prior}$ refers to the logarithm of the prior. Due to a high number of steps (100000) needed to describe our physical event the Gaussian distributions were assumed for the loglikelihood and the logprior. Then ℓ and $\wp_{prior}$ can be calculated as shown in equation (16) and (17). In equation (16), as in equation (14), d_i and m_i represent measurement data and model data at index i, respectively, and σ is the standard deviation of the measurement uncertainty. In equation (17) θ_j represents the j-th parameter in θ, and μ_j and σ_j represent the expected value and standard deviation, respectively, for the prior for this parameter. Equation (17) is only valid if the parameters of θ are independent. The final change that needs to be made is to alter step 7 to compare p_{move} to the value of $\ln(u)$ instead of u.

$$\ln\left(\frac{\mathcal{L}(d|\theta', M, I) p_{prior}(\theta'|M, I)}{\mathcal{L}(d|\theta^{i-1}, M, I) p_{prior}(\theta^{i-1}|M, I)}\right) =$$

$$\ell(d|\theta', M, I) + \wp_{prior}(\theta'|M, I)$$

$$-\ell(d|\theta^{i-1}, M, I) - \wp_{prior}(\theta^{i-1}|M, I)$$

$$(15)$$

$$\ell(d|\theta, M, I) = k + \sum_{i=1}^{n} -\frac{(d_i - m_i)^2}{2\sigma^2} \tag{16}$$

$$\wp_{prior}(\theta^{i-1}|M, I) = l + \sum_{j=1}^{m} -\frac{(\theta_j - \mu_j)^2}{2\sigma_j^2} \tag{17}$$

2.3.1 MCMC Setup

There are parameters that must be set for the MCMC algorithm, including the number of steps T, the standard deviation of the step σ_2. Additionally, the prior distributions for θ must be specified. Based on this the MCMC was run with the parameters and priors shown in Table 3. The prior probability distributions for C and h_{loss} are also shown in Figure 2.

Table 3. MCMC parameters and prior distributions

T	100000
σ_C	0.01
$\sigma_{h_{loss}}$	0.1
$p_{prior}(\theta|M, I)$	$p_{prior}(C|M, I) \cdot p_{prior}(h_{loss}|M, I)$
$p_{prior}(C|M, I)$	$\mathcal{N}(0.8, 0.05)$
$p_{prior}(h_{loss}|M, I)$	$\mathcal{N}(30, 5)$
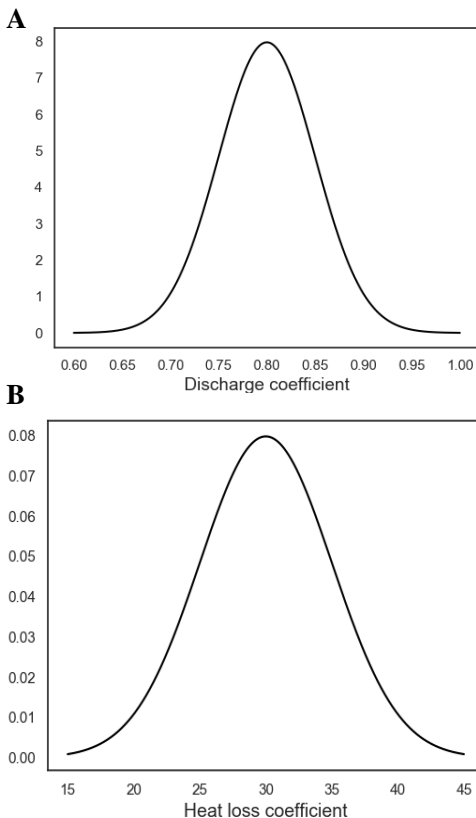
Figure 2. The prior probability distributions for A: discharge coefficient C and B: heat loss coefficient h_{loss}.

Additionally, some lead/lag compensation was implemented due to a time offset between the experimental and simulated data of approximately 4000 time-steps, consistent between experiments. It resulted

in a significant difference during the rise and fall of the pressure response in the experiments.

The model was solved using solve nonstiff differential equations- medium order method ode45 (MathWorks, 2020), with a maximum time step of 0.5 s.

2.3.2 Evaluation of MCMC representativeness

The representativeness and performance of the MCMC chains that were run were evaluated as described by Kruschke (2015). In this subchapter "the chains" refers to four chains run on experiment 11.

A visual inspection of the chains' trajectory shows that after a burn-in period, all the chains had converged to the same area. The first 3000 steps of each chain were then removed, considered to be part of burn-in. With these steps removed the chains were seen to overlap and mix well, and distinguishing the different chains from one another was basically impossible. This indicates that none of the chains are stuck with all the chains sampling from the same region. The convergence of chains with the starting position given by the X marker is presented in Figure 3.
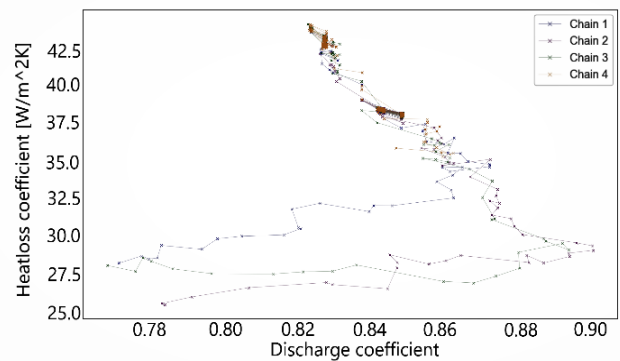
Figure 3. Convergence of 4 chains run in MCMC; exp 11.

The plots of marginal distributions were then created for the sampled parameter values for each chain, shown in Figure 4 and Figure 5. These density plots show high overlap, though with some difference in the peaks. Some difference in the density plot is to be expected, due to the finite number of samples drawn. The overlap between chains suggests the chains sample representative values from the posterior distribution.
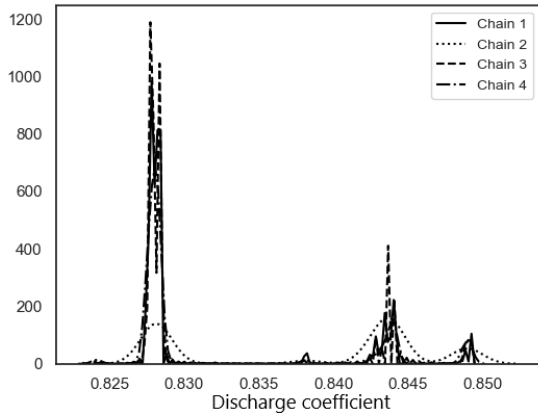
Figure 4. Density plot for parameter C for chain 1 through 4 for experiment 11.
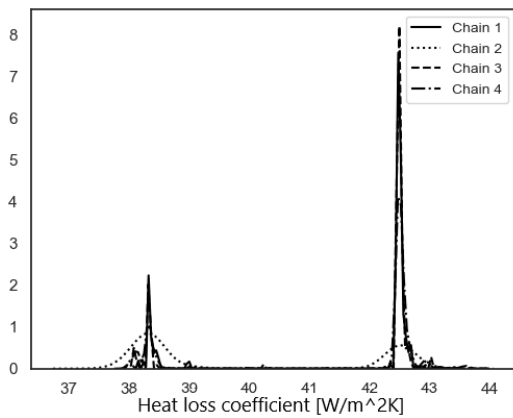
Figure 5. Density plot for parameter h_{loss} for chain 1 through 4 for experiment 11.

This evaluation of the MCMC gives us confidence in the results produced by the simulation. The evaluation shows no indication of poor representativeness nor low accuracy.

2.3.3 Posterior predictive check

If our model is adequate, then replication of the current experiment should generate data that are similar to the predictions made by the model. Let $y_t = y(t; \theta)$ denote the predicted pressure given $\theta = (C, h)$. We are interested in $p(y_t|D, m, I)$, where D and m denote the experimental data from the previous run and the model output (section 2.2), respectively. It follows from the marginalization and the product rules that

$$p(y_t|D, m, I) = \int_\Omega p(y_t, \theta|D, m, I) =$$

$$\int_\Omega p(y_t|D, \theta, m, I)p(\theta|D, m, I)d\theta \quad (18)$$

In case we know m and θ, it is reasonable to assume that previously acquired data and the prediction y_t are conditionally independent. This means that

$$p(y_t|D, \theta, m, I) = p(y_t|\theta, m, I) \quad (19)$$

Thus, the posterior predictive density is given by

$$p(y_t|D, m, I) = \int_\Omega p(y_t|\theta, m, I)p(\theta|D, m, I)d\theta \quad (20)$$

The second term of the integrand, $p(\theta|D, m, I)$ is the posterior density, which has been previously determined. The first term of the integrand, $p(y_t|\theta, m, I)$, is the model density which depends on the model and the measurement noise. The model density has the same functional form as the likelihood for a single data point. Note that, in the case of model density, this function is a probability density with respect to y_t, with the parameter θ assumed to be known.

The expression for the posterior predictive density (20) shows that the uncertainty in the predicted pressure is due to two processes. The first one is contributed by the posterior density, which expresses the uncertainty about the true value of θ. The second contribution is due to the measurement noise. Although, more experimental evidence can reduce the uncertainty about θ, nevertheless, the measurement noise cannot be eliminated. In the maximum likelihood approach, it is assumed that the inferred value θ_{ML} for the unknown parameter is the true value and therefore the only source of uncertainty is the measurement noise. Indeed, in the context of eq. (20), in the maximum likelihood approach one claims that

$$p(\theta|D, m, I) = \delta(\theta - \theta_{ML}) \quad (21)$$

and hence, the integral in (20) reduces to

$$p(y_t|D, m, I) = p(y_t|\theta_{ML}, m, I) \quad (22)$$

As the evidence grows, we will become more confident about the true value of θ and hence the posterior density becomes narrower and in the limit it converges to (21). Nevertheless, as long as the uncertainty about the true value of θ is large, the Bayesian and maximum likelihood approaches will differ.

Due to difficulties to find an analytical expression for y_t, in general, a Monte Carlo based approach is more viable. The algorithm below describes the steps.

Algorithm 1

Generating y_t; for $i = 1, ..., N$ repeat the following steps

1.	$\theta_i \sim p(\theta|D, m, I)$
2.	$y_i \sim p(y|\theta_i, m, I)$

Applying this algorithm for large N (equal to the number of MCMC samples), the empirical distribution of the samples y_i will approximate $p(y_t|D, m, I)$.

3 Results and discussion

Hydrogen jet fires are causing high overpressure due to the high amount of released energy in a very short time. The molar balance and temperature in the enclosure described in the methodology explained the pressure dynamic.

In the model, the mass balance is crucial for the accuracy of overpressure prediction. Therefore the discharge coefficient used in the calculation of molar flow through the vents (in and out) needs to be

investigated. A discharge coefficient is a dimensional number representing flow and pressure loss through the orifice. Is a function of Reynold number $C = f(Re)$ while the Re is a function of the flow rate, hence the $E_{loss} \sim u^2$. Some assumptions have been used for different scenarios but when the discharge coefficient is unknown or uncertain it has to be found experimentally. For the computing methods, it is recommended to assume C=1 (Crowl and Louvar, 2011). From the previous study on PPP it is known that C=1 is an assumption that can't be applied, and lower values showed better accuracy. The reasoning of choosing the values of C in the work of Hussein et al., (2018) and Makarov et al., (2018) was based on the literature knowledge and validated against their experiments. The values presented in their work are not. It can be due to different experimental setups and mass flow rates. In the mentioned studies the heat transfer shows its importance for the pressure dynamics. Nevertheless, the heat loss has been neglected in the model (due to the small scale of experiments).

In the model presented in this paper a simple solution of the heat loss is used. The water condensation hasn't been included in the numerical simulations. The heat loss described in section 2.2 determines the rate of heat transfer through the walls.

The MCMC evaluation results in the posterior distributions. Experiment 11 has been chosen to presents the parameter estimation analysis. In the figure below (Figure 6) the posterior pdf is presented. The clusters you can see on the plots represent the area of 'the best' θ. Experiment 11 resulted in bimodal posterior distributions. By looking into the density plots of C and h_loss (Figure 4 and Figure 5) the two major peaks are clearly visible. Two clusters represent two areas of θ value which are consistent with the experimental results.
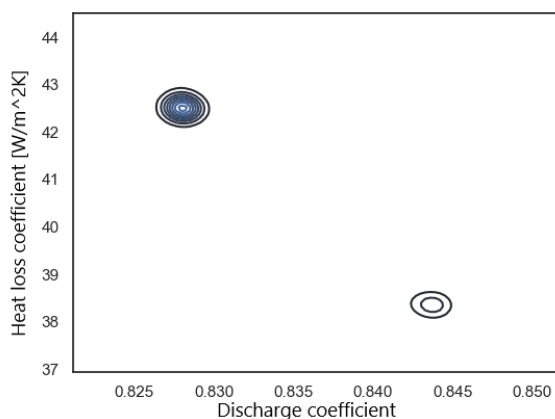
Figure 6. Bivariate posterior distributions for experiment 11.

Due to uncertainty in measurements, in general, it is not possible to uniquely determine the values of the unknown parameters. In the Bayesian approach, this problem is addressed by the posterior distribution. The posterior distribution summarizes the belief in the probable range of the values for the unknown parameters (Kruschke, 2015). Models are simplified versions of reality. In this context, the simplification means that we only take into account certain aspects of the reality and assume that the other aspects either cancel out each other or have no significant influence on the description or predicted behaviour of the system of interest. One way to evaluate how good these simplifications are is to check how accurately the model mimics the data. Therefore, it is important to not confuse the model inadequacies with the uncertainty due to inference. Thus, it is important to keep track of the sources of uncertainty. One way to do this is by conducting the so called posterior predictive check.

The MCMC results have been applied into the PPP model (Figure 7). Both experimental and simulation results have their uncertainty, included in curves. For experimental results, the uncertainty is ±1% FSO BFSL (grey area around black line Figure 1). For the simulation, the uncertainty was calculated at the specific time for each sample (vertical red lines from blue dots Figure 7). The simulation overestimates the actual pressure with 0.6 kPa in maximum pressure and underestimates in minimum pressure is 0.6 kPa as well which is acceptable.
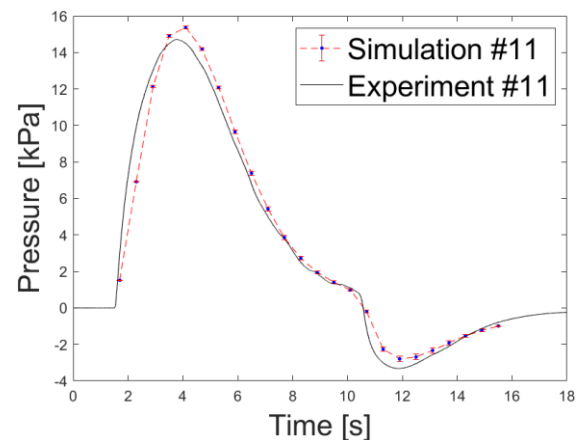
Figure 7. Pressure dynamics in 14,9 m³ the enclosure during experiment 11: experimental (black line), simulation with estimated θ individualy for each time step (blue dots). The standard deviation is represented by error bars (red lines with caps).

The simulation result shows good accuracy with experiments (Figure 8).
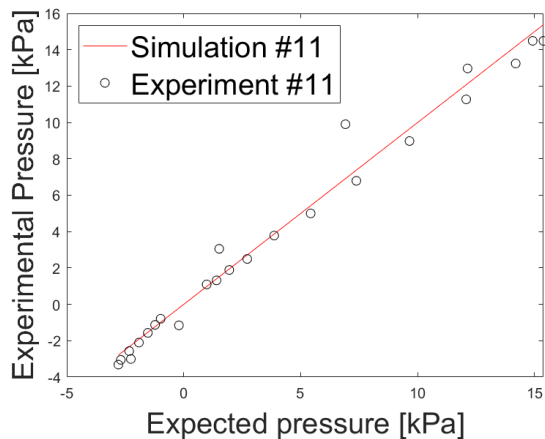
Figure 8. Exp11 pressure predicted (red line) vs. measured (black circles).

4 Summary

In this paper, we have used a Bayesian approach to estimate parameters in a PPP model. The intended use for this model is to simulate overpressures from accidental hydrogen releases in confined spaces. The parameters of interest were the heat loss coefficient and the discharge coefficient. The data used in creating the model, and used when estimating the parameters, was collected from large scale experiments performed by the University of South-Eastern Norway.

Markov Chain Monte Carlo was used to generate samples from the posterior distribution. The performance of the MCMC algorithm was evaluated, and seen to perform well.

In this analysis, we found that a discharge coefficient of $C = 0.9$ and a heat loss coefficient of $h_{loss} = 30$ are the most likely values which capture the results across all the experiments. The Bayesian analysis of the model gave the most probable values for performed experiments (set up dimension and flow rates).

References

S. Brennan, D. Makarovand, and V.V. Molkov. Dynamics of Flammable Hydrogen-Airixture Formation in an Enclosure with a Single Vent. *In Proceedings -Fire and Explosion Hazards 6, 11.-16. April, Leeds, UK*, 493–503, 2010. doi: 10.3850/978-981-08-7724-8_07-04.

S. Brennan and V.V. Molkov. Pressure peaking phenomenon for indoor hydrogen releases. *International Journal of Hydrogen Energy,* 43(39), 18530–18541,2018. doi: 10.1016/J.IJHYDENE.2018.08.096

D.A. Crowl and J.F. Louvar. *Chemical Process Safety Fundamentals with Applications*, 3rd ed.; In B. Goodwin, J. Fuller and E. Ryan, editor. Prentice Hall 2011.

F. Daly, D. Hand, M. Jones, A. Lunn, and K. McConway. *Elements of Statistics (1st ed.).* Addison Wesley, The Open University, 1995.

H.G. Hussein, S. Brennan, V. Shentsov, D. Makarov, and V.V Molkov. Numerical validation of pressure peaking from an ignited hydrogen release in a laboratory-scale enclosure and application to a garage scenario. *International Journal of Hydrogen Energy*, 43(37), 17954–17968, 2018. doi: 10.1016/J.IJHYDENE.2018.07.154.

J.K. Kruschke. Doing Bayesian Data Analysis, A tutorial with R, JAGS and Stan. *In Becoming a Brilliant Trainer (2nd ed.) 2015*. doi: 10.4324/9781315627960-16.

A.W. Lach. Pressure Peaking Phenomena-unignited releases. Internal *Report experiments, 2019*.

D. Makarov,V. Shentsov, M. Kuznetsov, and V.V Molkov. Pressure peaking phenomenon: Model validation against unignited release and jet fire experiments. *International Journal of Hydrogen Energy*, 43(19), 9454–9469,2018. doi: 10.1016/J.IJHYDENE.2018.03.162.

MathWorks. Matlab User Guide(r2020b). *ode45* Apple Hill Drive Natick, MA, 2020.

B.J. Mcbride, S. Gordon, and M. Reno. *Coefficients for Calculating Thermodynamic and Transport Properties of Individual Species.* Nasa Technical Memorandum, 4513(NASA-TM-4513), 98, NASA, US. Government 1993.

J. A. Vrugt. Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environmental Modelling & Software*, 75, 273–316, 2016. doi: 10.1016/J.ENVSOFT.2015.08.013.

J. Wang, Z. Zhou, K. Lin, C.K. Law, and B. Yang. Facilitating Bayesian analysis of combustion kinetic models with artificial neural network. *Combustion and Flame*, 213, 87–97,2020. doi: 10.1016/j.combustflame.2019.11.035.

Classification of Gases and Estimation of Gas Flow Rate Based on Unsupervised and Supervised Learning Respectively

Maths Halstensen[1]   Simen S. Halstensen[2]

[1]Department of Electrical Engineering, IT and Cybernetics, University of South-Eastern Norway, maths.halstensen@usn.no
[2]Department of Electrical Engineering, IT and Cybernetics, University of South-Eastern Norway, simenshalstensen@gmail.com

Abstract

In this research project, acoustic chemometrics was assessed as a new method for both classification and prediction of flow rate of five selected gas types. The gas types were selected to span different densities as much as possible while at the same time being relatively safe to use. The five gas types were Argon, Helium, Carbon dioxide, Nitrogen and Air. The research questions were 1) Can measurements of the vibrations in a gas control valve in combination with signal processing and unsupervised learning be used to classify the five gases mentioned above? 2) Can the vibrations in the gas valve in combination with supervised learning be used to determine the flow rate of the five gases? 3) Can a simple low cost piezo disk provide signals comparable to that of an industrial accelerometer? The results show that it is possible to classify the five gas types based on principal component analysis with three components. The gas flow rate could also be predicted for all five gases based on partial least squares regression with an average error of 2-5%. The Piezo disk could not be used for gas classification, but for prediction of gas flow rates it was comparable to the accelerometer. All the prediction models were validated based on independent data.

Keywords: *supervised learning, unsupervised learning, partial least squares regression, classification, principal component analysis, accelerometer*

1   Introduction

Acoustic chemometrics (Halstensen, 2006, 2010; Esbensen, 1999; Arvoh, 2012) is a relatively new general process analytic approach for on-line monitoring of industrial processes. Acoustic chemometrics concerns capturing characteristic system vibrations generated by e.g. a transportation or manufacturing process.

The vibrations -or acoustic signatures -can easily be measured by a non-intrusive sensor (accelerometer).

Previous research has shown that these acoustic signatures carry embedded information about physical and chemical parameters, such as composition (oil, fat, ammonia, buttermilk, glycol, and ethanol), mixing progress, fiber length, flow, density, temperature - as well as system state. For extraction and quantification of these types of specific analytes and parameters of interest, domain transforms (Fast Fourier Transform) (Ifeachor & Jervis, 1993) is essential prior to multivariate analysis and modelling.

Unsupervised methods such as Principal Component Analysis (PCA) (Esbensen and Swarbrick, 2018) can be used to e.g. classify products. Supervised methods such as Partial Least Squares (PLS) (Martens & Næs, 1989) regression based on empirical input-response data are used to extract relevant information and to calibrate multivariate regression models. The models can then be used either to classify or to predict the parameter of interest based on mew independent acoustic spectra. As with all multivariate models, the role of proper validation is critical (Esbensen and Geladi, 2010).

Advantages of acoustic chemometrics:
1. clamp on non-intrusive sensor
2. No process modifications necessary
3. Several predictions from the same measurement

The fact that almost all processes produce some kind of acoustic emission opens up for many potential applications, all of which depend on multivariate calibration. The only requirement is that the acoustic emission must contain relevant information, which can be correlated with the parameter(s) of interest. In this work, we test the feasibility of acoustic chemometrics to both classify gas type and predict gas flow rate. The method is demonstrated on five different gas types: Argon, Helium, Carbon dioxide, Nitrogen and Air, which is a mixture.

The proposed method cannot be directly compared to the accuracy of e.g. Coriolis type flow meters (LaNasa & Upp, 2014), but the advantage of this method is that it is non-intrusive, inexpensive, easy to implement on exiting equipment (no interference with existing

equipment is necessary). The industry often prefers non-invasive or non-intrusive measurement methods since parts sticking into the process pipelines can cause critical fouling and clogging which is problematic but also leads to faulty measurements.

The main objective in this research project is to test a new method for gas flow metering and show that it works on different gas types. The second objective is to show that the same method also can be used to classify the gas type directly from the same raw measurement as is used for the flow rate estimation. The last objective is to validate the method based on independent experimental data to ensure optimal model complexity and estimation of the uncertainty of the method.

Very little causal interpretation of the acoustic spectra is reported in literature. In a study from 2008 (Kupyna et al., 2008] tried to explain the physical theory behind the flow of liquids through an orifice plate and the vibrations that occurred. The experimental results agreed with theory that they found, but it is difficult to see how it can be directly used to improve acoustic chemometrics. We have not attempted to explain the physical theory behind the vibrations that occur in the valve from the different flowing gases in this study.

2 Materials and methods

The experiments were carried out on an experimental gas flow rig in the chemometrics laboratory at University of South-Eastern Norway, Porsgrunn, Norway.

2.1 The experimental gas flow rig
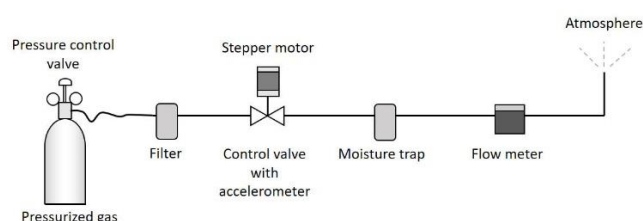
Figure 1. Experimental gas flow rig at USN.

All the experiments with gas flow were carried out in the chemometrics laboratory at the University of South-Eastern Norway, Porsgrunn, Norway (USN).

The experimental rig used to develop the method is shown in Figure 1 above. The rig consists of a pressurized gas bottle, a filter, a gas control valve, a moisture trap and a gas flow meter. All the elements are connected with hoses and pipes, which were pressure-tested to avoid gas leaks in the system prior to the gas flow experiments. The filter is there to avoid getting particles into the control valve. These particles can come from connecting and disconnecting the hoses and clamps at around the gas bottle.

The gas control valve is a simple rotary gas valve used in compressed air systems. A stepper motor is

connected directly to the valve to be able to control the gas flow rate automatically. The control loop was implemented in LABVIEW 2017 from National Instruments. A large moisture trap/filter normally used for spray painting is located right in front of the gas flow meter to avoid moisture entering the flowmeter. Moisture can damage the gas flow meter or cause false readings.

The gas flow meter Mass-Stream D-6300 from BronkhorstTM has a gas flow range up to 200 l/min of air. Since we here have other gas types a conversion factor was necessary to get the correct flow rates for all other gases than air. The accuracy of the flow meter is +/- 1 %.

The gas leaving the gas flow meter was injected into the ventilation system and finally disposed to the atmosphere. All the elements mentioned above was connected with hose with inner diameter = 10mm. On the gas control valve shown in Figure 2 the two sensors (accelerometer and piezo disk) were glued on to the main metal body of the valve.

Figure 2. Gas control valve with the stepper motor connected from top.

2.2 Gas types

Five different gases were selected: air, argon, carbon dioxide, nitrogen and helium. These gases were selected because we wanted to span the densities as much as possible but at the same time, we wanted to use gases, which are relatively safe. Air is a mixture containing mostly nitrogen (78%) and oxygen (21%) so in the gas classification it should behave similar to nitrogen.

2.3 Acoustic sensors S1 and S2

S1 is an industrial standard accelerometer with a frequency range of 0-40kHz, S2 is a simple 0.5mm thick piezo electric element. Figure 3 shows a picture of the two vibrational sensors S1 and S2.

Figure 3. Vibrational sensors, S1 (left) is an accelerometer from Endevco 7259B, S2 (right) is a simple low cost ($1) piezo electric disk with resonant frequency 4 kHz.

Two different sensors were mounted in order to test if the results from a low cost piezo element (S2) would be comparable to the results based on the much more expensive accelerometer (S1).

The measurement systems also consist of a Signal Amplifying Module (SAM) as shown in figure 4 and a data acquisition system from National Instruments (USB-6351) connected to a computer. The SAM unit is a special experimental equipment developed and built at USN.

Figure 4. Signal Amplifying Module (SAM)

The signal from the accelerometer was amplified 100 times, while the signal from the piezo disk was only amplified 40 times.

2.4 Signal processing

The signals from the sensors S1 and S2 were converted from time to frequency domain using Fourier transformation (Ifeachor & Jervis, 1993) prior to analysis and model calibration. The Discrete Fourier transform (DFT) can be expressed as

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi kn/N} \quad k = 0, \dots, N-1 \qquad (1)$$

A more efficient implementation of the DFT is the Fast Fourier Transform (FFT), which in this work has been implemented in LabVIEW 2017 for fast real time calculation of the Fourier spectrum.

Each measurement, which was a time series of 4096 samples recorded with a sampling frequency of 300 kHz, were transformed using Fast Fourier Transform. The resulting spectrum contains 2048 frequencies.

2.5 Principal Component Analysis

Principal component analysis (PCA) is a standard multivariate data analysis method (Esbensen and Swarbrick, 2018), PCA uses an orthogonal transformation to convert a set of observations often described by many correlated variables into a few linearly uncorrelated latent variables called principal components. Principal Component Analysis (PCA) which is a so-called unsupervised method can be used for classification of observations. It is called unsupervised because of the lack of reference data to guide the decomposition. The decomposition of a multivariate matrix X is shown in equation 2.

$$X = T\,P^T + E \qquad (2)$$

The T matrix contains the score vectors, P contains the loading vectors and E is the residual.

In this work, the NIPALS algorithm (Ergon et al., 2009) was used because of one of its advantages, which is that it is possible to calculate only a relevant number of components. Principal component analysis can alternatively be carried out based on singular value decomposition (SVD). The advantage of using SVD is that the code is simple, the disadvantage is that all possible components always is calculated so if X is large (I.e. is made up of a high number of variables), SVD is time consuming. NIPALS also works on matrices with missing values, in which case SVD will fail.

2.6 Partial Least Squares Regression

Partial least squares regression PLS-R was used to calibrate multivariate prediction models for flow rate of all five gas types. The reference measurements from the flow meter in the experimental flow loop was used as the reference during calibration.

Partial Least Squares Regression (PLS-R) is a supervised modelling approach which is well explained in literature (Martens & Næs, 1989; Ergon et al., 2009) thus only a short introduction is given here.

PLS-R relies on representative training data for two variable blocks, often called X and Y respectively. In the present study, the X data matrix contains the acoustic frequency spectra, and Y is a vector containing the gas flow rates from the gas flow meter used as a reference. The NIPALS algorithm is the most widely used algorithm in PLS regression. In this algorithm, the intention is to model both X and y simultaneously, make the error as small as possible and at the same time extract as much useful information from the X matrix in order to describe the y response variable. A simplified version of the NIPALS algorithm is presented below (Ergon et al., 2009).

A is the optimal number of components in the model.
1. Let $X_0 = X$. For $a = 1, 2, \ldots, A$ perform steps 2 to 6
2. $w_a = X_{a-1}^T y / \|X_{a-1}^T y\|$ (with length 1)
3. $t_a = X_{a-1} w_a$
4. $q_a = y^T t_a (t_a^T t_a)^{-1}$
5. $p_a = X_{a-1}^T t_a (t_a^T t_a)^{-1}$
6. Compute the residual $X_a = X_{a-1} - t_a p_a^T$

$$X = T_w P^T W W^T + E \qquad (3)$$
$$y = T_w q_w + f \qquad (4)$$

where the score matrix $T_w = [t_1 \quad t_2 \cdots \quad t_A]$ is orthogonal, loadings matrix $P = [p_1 \quad p_2 \cdots \quad p_A]$, $q_w = [q_1 \quad q_2 \cdots \quad q_A]$ and the loading weight matrix $W = [w_1 \quad w_2 \cdots \quad w_A]$

The loading matrix, P, is calculated as
$$P = X^T T (T^T T)^{-1} \qquad (5)$$

The prediction vector for $y = Xb + f$ corresponds to:
$$\hat{b} = W(W^T X^T X W)^{-1} W^T X^T y \qquad (6)$$

The response vector $\hat{y} = X\hat{b}$ (7)

In evaluating the regression model, the root mean squared error of prediction RMSEP offset, slope and correlation coefficient are commonly used. The root mean squared error of prediction is calculated as:

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_{i,\text{predicted}} - y_{i,\text{reference}})^2}{n}} \qquad (8)$$

3   Experimental

The gas bottle was connected to a constant pressure valve, which were connected to the experimental gas flow rig. The constant pressure valve was adjusted to four bar. The set point of the gas control valve was entered and the gas started to flow. When the gas flow rate became stabile after a few seconds, the acoustic measurement was started. The gas flow rate was simultaneously measured and recorded by the flow meter in figure 1. The gas flow rate levels were chosen randomly until all the 21 discrete levels from 1 to 100 l/min were covered. The discrete gas flow levels were 1, 5, 10, 15, 20, 25,…, 100 [l/min]. The experiment was repeated following the same strategy as in the first round to get independent data for validation of the models. This was the experimental procedure followed for all the five gas types.

In order for the gas flow meter to give reliable flow rate measurements for the gases different from air, a set of conversion factors were necessary to scale the signal to the correct level. The conversion factors were provided by the manufacturer of the flow meter. The conversion factors used can be seen in table 1.

Table 1. Gas flow rate conversion factors

Conversion factors for the five gas types measured by the flowmeter:	
Air:	1.00
Argon (Ar):	2.01
Carbon dioxide (CO2):	1.20
Helium (He):	0.24
Nitrogen (N2):	1.00

Seven replicates of acoustic spectra were recorded for each gas and flow rate. Each of these replicates was an average of 200 individual spectra. The averaging is necessary to reduce white noise and get a precise measurement. Each batch of seven replicates took approximately 35 seconds to record.

During the experiments on CO2, a significant amount of ice was building up around the valve on the gas bottle. This is normally not a big problem but in our case, it was problematic because we used hoses to transport the gases to the flow rig and these hoses became quite fragile at this low temperature. Fortunately, no hoses broke during the experiments.

4   Results and discussion

4.1   Classification of gas types (S1)

The score plot t1-t2 in Figure 5 shows that principal component 1 describes gas flow rate for all the five gas types, the gas flow rate increases from right to left in the score plot. The loading vector (Figure 5) for principal component 1 which indicates that all the frequencies in the acoustic frequency spectra increase with increasing gas flow rate since all frequencies have the similar loading value. All the results in Figure 5 is based on data from the accelerometer only.
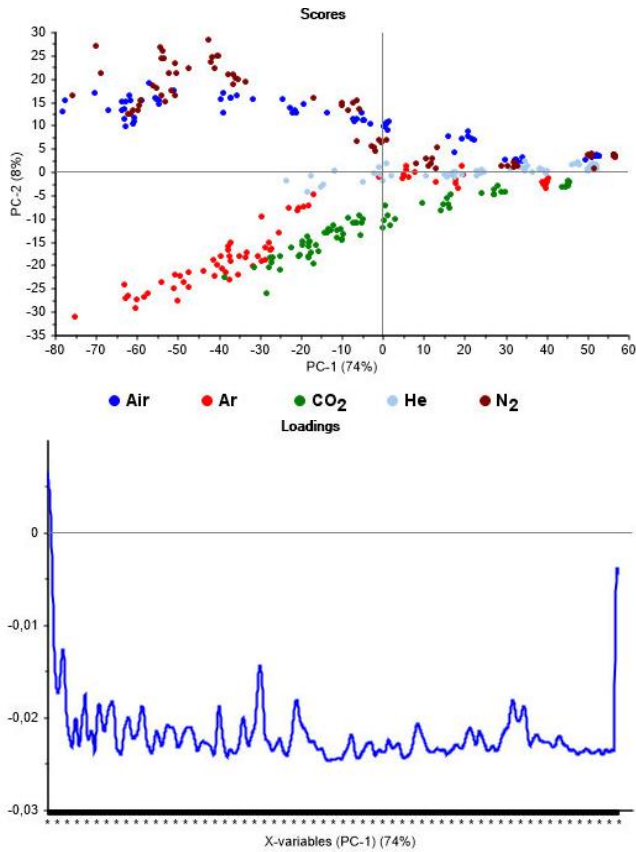
Figure 5. Upper: Score plot t1-t2, Lower: loading vector p1.

Figure 6 shows the classification of the five different gas types. The score plot in Figure 6 shows all the five gas types with different colors.

The score plot in Figure 6 above shows that the gas groups/clusters are located in a star pattern. It is obvious why it has to be exactly like this because "no flow" sounds exactly the same for any gas, namely no sound at all! The zero flow (0 l/min) is therefore the center of the "classification star". As expected, the nitrogen (N2) group overlaps significantly with the cluster for air because air contains 78% of nitrogen.

The Loadings for the first two components shows that there are relevant information in almost all the frequencies in the spectra.

The explained variance plot shows that most of the variance in the data is caused by variation in airflow rate as PC-1 explains 78% of the variance. PC-2 and PC-3, which is used for classification of the gases, only explains 8% and 4% respectively.

The PCA classification model was based on data from two independent experiments separated in time. An alternative strategy would be to determine the PCA classification model based on one of the experiments and used the other one for testing. Based on the results shown in the score plot in Figure 6 we have here only shown that classification of these five gases based on

vibration analysis is promising. Further validation of the concept is needed to verify reliability over time.
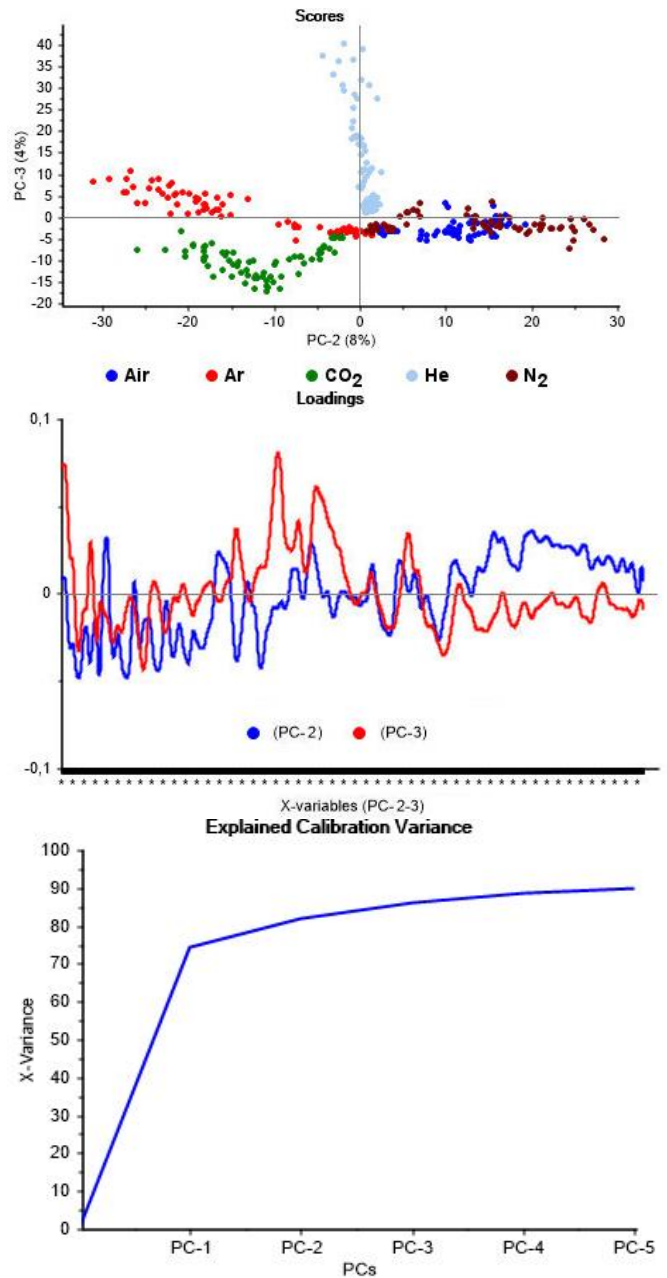
Figure 6. Classification of gases based on the accelerometer data, Upper: Score plot t2-t3, Middle: Loading vectors p2 and p3. Lower: Explained variance.

4.2 Classification of gas types (S2)

From the analysis of the data from the low cost piezo disk it became clear that it was not possible to completely separate and thus classify the five gases. Figure 7 shows the score plot t2-t3 of the piezo disk data, which is directly comparable to the score plot in figure 6 representing the accelerometer. As can be seen in figure 7 the data is more spread (noisy) leading to overlap between gas types. The PCA classification model was also here based on data from two

independent experiments separated in time as described in Section 4.1
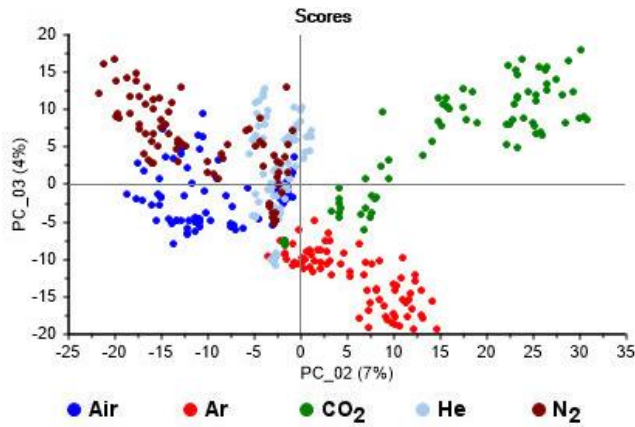
Figure 7. Classification of gases based on the piezo disk data, Score plot t2-t3.

4.3 Estimation of gas flow rate
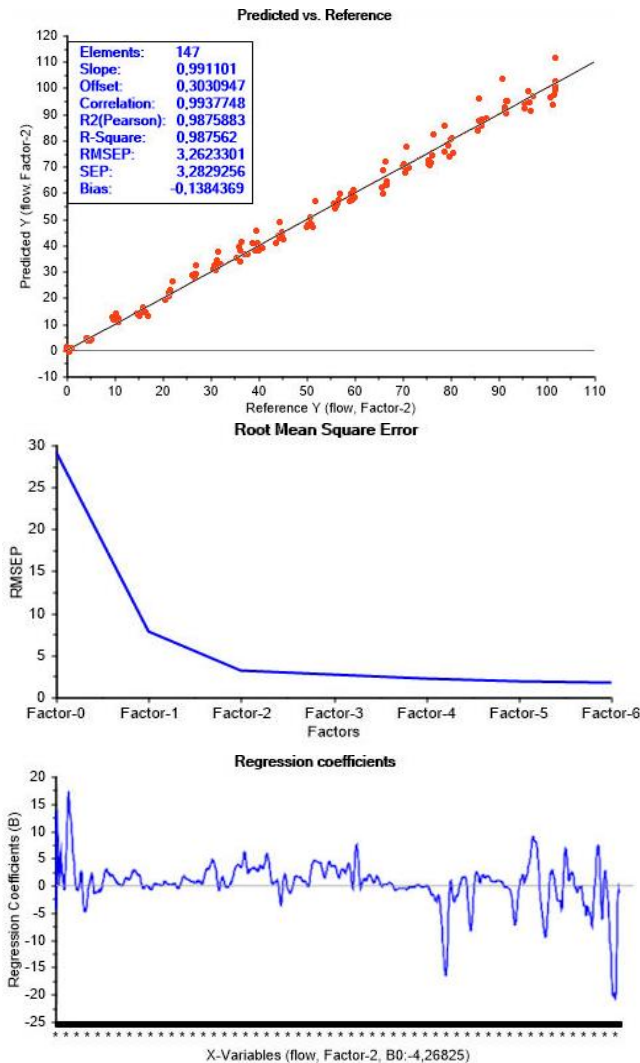
Figure 8. PLS regression model for prediction of gas flow rate of Argon, 2 components was optimal and no outliers were deleted.

Five PLS-R models were calibrated for flow rate of the five different gas types using the acoustic data (spectra) as the independent X-matrix and the reference Y-vector was the flow rates from the gas flow meter.

The prediction performance of the five models were similar with comparable RMSEP values. Only the full model result for Argon is shown in Figure 8 while the results for all the 5 gas types has been summarized in table 2 for comparison. The model for argon flow rate in figure 8 is based on both sensors (Accelerometer and Piezo disk) combined.

Table 2 below shows a comparison of all the models for gas flowrate of all five gases, and all combinations of sensors.

Table 2. Comparison of gas flow rate models for Air, N2, CO2, He and Argon. The lowest RMSEP values of the models based on single sensors are indicated with bold numbers.

Gas type	Both sensors (4096 variables)			Accelerometer (2048 variables)			Piezo disk (2048 variables)		
	RMSEC	RMSEP	#Comp.	RMSEC	RMSEP	#Comp.	RMSEC	RMSEP	#Comp.
Argon	3.1583	3.2623	2	3.5545	3.8124	2	3.5275	**3.4026**	2
Air	2.8109	4.0350	2	3.1184	**4.1794**	2	3.0684	4.2792	2
CO2	3.3641	3.3641	3	3.8272	5.0107	3	3.8713	**4.6002**	3
He	2.2870	2.5896	4	1.6909	**2.0514**	3	4.8856	4.9125	4
N2	1.8420	3.5744	2	2.3368	3.6871	2	1.7049	**3.6106**	2

A comparison of the RMSEP values of the models based on single sensors shows that the sensors performs very similar and both can be used for prediction of gas flow rate. However, if classification of gas type is of interest it is only the accelerometer, which can be used. The loading plots in Figure 5 and 6 and the regression coefficients in Figure 8 shows that there is information in all frequencies in the acoustic spectra, an attempt to use variable selection where some of the variables are omitted in the model will fail because of this. In order to improve the classification or prediction of gas flowrate, the most effective method is to use moving average either in the frequency direction or in the time direction. Filtering in the time direction will often result in a lower RMSEP but since we introduce a low-pass filter, the model will not be able to detect rapid changes of the parameter of interest.

5 Conclusion

A new method for non-intrusive gas classification and flow rate estimation is proposed. The research objectives was 1) assess acoustic chemometrics potential to classify 5 selected gases, 2) predict the flow rate of each gas in the range 0-100 l/min and 3) compare results from a standard accelerometer and a simple low cost piezo disk. The results in section 4 shows a successful classification of the gas types based on PCA with three components on the measurements from the accelerometer. The results also show that it is possible to predict gas flowrate based on vibrations in the gas

flow rate valve. The RMSEP values shows an average error between 2 and 5% for the gas flow rate models when tested on independent data. From the comparison of the low cost piezo disk and the accelerometer, it was clear that the piezo disk could not be used for gas classification, but for prediction of gas flow rates it was comparable to the accelerometer.

References

B. K. Arvoh, S. Asdahl, K. Rabe, R. Ergon, and M. Halstensen. Online estimation of reject gas and liquid flow rates in compact flotation units for produced water treatment. *Flow Meas. Instrum.*, 24: 63–70, 2012. doi:10.1016/j.flowmeasinst.2012.03.008

R. Ergon. Re-interpretation of NIPALS results solves PLSR inconsistency problems. *J Chemometrics,* 23: 72-75, 2009. doi:10.1002/cem.1180

K. H. Esbensen and B. Swarbrick, Multivariate Data *Analysis – An introduction to Multivariate Analysis, Process Analytical Technology and Quality by design*, 6th Ed., *CAMO Publishing*, 2018. ISBN 078-82-691104-0-1

K. H. Esbensen and P. Geladi. Principles of Proper Validation: use and abuse of re-sampling for validation. *Journal of Chemometrics,* 24: 168-187, 2010. DOI: 10.1002/cem.1310

K. H. Esbensen, B. Hope, T. T. Lied, M. Halstensen, T. Gravermoen, and K. Sundberg. Acoustic chemometrics for fluid flow quantifications: II. A small constriction will go a long way. *Journal of Chemometrics,* 13 (27): 209–236, 1999. doi: 10.1002/ (SICI) 1099-128X (199905/08)13:3/4b209::AID-CEM553>3.0.CO;2–5.

M. Halstensen and K. H. Esbensen. Acoustic chemometric monitoring of industrial production processes in: K.A. Bakeev, *Process Analytical Technology,* Wiley, Chichester, United Kingdom, 2, 281–302, 2010. doi:10.1002/9780470689592, ISBN: 978-0-470-72207-7

M. Halstensen, P. de Bakker, and K. H. Esbensen. Acoustic chemometric monitoring of an industrial granulation production process—a PAT feasibility study. *Chemometr. Intell. Lab Syst.,* 84: 88–97, 2006. doi:10.1016/j.chemolab.2006.05.012

E. Ifeachor and B. Jervis, *Digital Signal Processing, a Practical Approach*, *Addison-Wesley Publishing,* 1993. ISBN-13: 978-0201596199

A. Kupyna, R. B. Schüller, E-O. Rukke, and T. Isaksson. Acoustic chemometrics on liquid flow: Shift in the frequency spectra and its relationship to the physical properties of the liquid and the pipe. *Chemometr. Intell. Lab. Syst.,* 91 (2): 151-163, 2008. doi:10.1016/j.chemolab.2007.10.010

E. L. Upp and P. J. LaNasa. *Fluid Flow Measurement: A Practical Guide to Accurate Flow Measurement. Gulf Professional Publishing,* 2002. **ISBN:** 9780884157588

H. Martens and T. Næs. *Multivariate Calibration*, *Wiley*, Chichester, UK, 1989 reprint 1994. ISBN-10: 0471930474, ISBN-13: 978-0471930471

Partial least squares PLS1 vs. PLS2 – optimal input/output modeling in a compound industrial drying oven

Maths Halstensen[1]   Ulrich Hundhausen[2]   Kim H. Esbensen[3]

[1]Department of Electrical Engineering, IT and Cybernetics, University of South-Eastern Norway,
Norway, maths.halstensen@usn.no
[2]Norwegian Institute of Wood Technology (NTI), Oslo, Norway, uhun@treteknisk.no
[3]KHE Consulting, Copenhagen, Denmark, khe.consult@gmail.com

Abstract

A feasibility study was carried out to assess the possibility of developing prediction models for monitoring drying conditions of wood coatings in one of Europe's largest and most modern coating plants for exterior cladding. These models were based on data from real-time Process Analytical Technology (PAT) sensors, measuring airflow and air direction, temperature and relative humidity). The study revealed that the information from the PAT sensors gave sufficient input to accurately model the complex drying conditions and their interrelations. Modelling was carried out using both Principal Component Analysis (PCA) and PLS-regression in both its PLS1 and PLS2 manifestations. In addition, the diagnostic prediction performance RMSEP between PLS1 and PLS2 models were not significantly different. This is advantageous for an industrial implementation concerning re-calibration operations: PLS1 requires 40 separate calibrations whereas PLS2 requires only one, because PLS1-R is a regression of a singular output variable (y-variable) and PLS2-R of several simultaneous, correlated output variables. While a single calibration based on PLS2 will take approximately one hour, the PLS1 approach will take more than a week.

Keywords: process monitoring; exterior wood coatings; wood cladding; oven drying; multivariate modeling; PLS1-R, PLS2-R

1 Introduction

Wooden cladding is the most common façade material for residential buildings in Norway, most often painted with multi-coat opaque high-build (>60-100 µm dry film thickness) to very high-build systems (>100 µm dry film thickness). Traditionally cladding has been painted by hand after installation, but the last years have seen a strong trend towards industrial coating application in Scandinavia (Hundhausen et al., 2016). Industrial application has the advantage of offering customers cladding that does not need immediate maintenance after mounting; moreover, industrial coating takes place under controlled conditions, which is of special importance as water-based coating systems today dominate the marked driven by environmental requirements for reduced amounts of volatile organic compounds (Elliott and Glass, 2000). While in the furniture and joinery industry UV-curable coatings are widely used, wood products for exterior applications are finished as physically dried coatings. Thus air velocity, relative humidity (RH) and ambient air temperature are the three decisive factors in convection drying (Nienhuis, 2014) causing water to flow from the lower to the upper part of coating films (Trent, 1992; Beetsma, 1996; Vanderhoff, 1966). Deficient drying can cause *blocking*, i.e. coated parts stick together after stacking which results in considerable transport damages, or can significantly affect coating adhesion, which promotes coating defects like paint *flaking*. In addition, the dry-to-recoat stage in industrial drying is critical, as it is one of the major factors controlling the overall batch production rate. Dry-to-recoat is the time interval needed before a second coat can be applied without risks for further defects, such as blistering or loss of adhesion (Koleske, 2012). In short, controlling drying is of overall importance regarding process capacity and yield.

The Norwegian cladding producer Gausdal Bruvoll SA has recently established one of the largest and most modern coating plants for industrial cladding in Europe (Hundhausen et al., 2018a). Controlled fast drying is crucial but technically challenging as different "orders" (customer specified batches of boards) often with different coatings pass through the approximately 70 m long multi-level dryer with a feed speed of 18 m/min (Hundhausen et al., 2016, 2018b). Reliable on-line drying monitoring is highly desirable to increase production yields by avoiding *coatings defects*.

A feasibility study has been carried out to assess the possibility of developing prediction models for monitoring drying conditions of exterior wood coatings in 8 out of 9 possible trajectories in Gausdal Bruvoll's dryer. Can drying conditions air flow, transport direction, air temperature and relative humidity recorded through a set of real-time Process Analytical Technology (PAT) (Bakeev, 2010) sensors be acquired accurately enough for process operators to run the drying facility reliably with a view of process optimality across many types of paints and coating combinations?

The first research objective in the present study is to assess the real-time on-line information from a set of

strategically placed PAT sensors as input variables for modeling.

The second research objective is strictly data analytical. Partial Least Squares 1 (PLS1) regression modelling is compared with its PLS2 alternative (Marten and Næs, 1989; Esbensen and Swarbrick, 2018). PLS1-R is regression of a singular output variable (y-variable); PLS2-R is regression of several simultaneous, correlated output variables (y-variables). The rationale for evaluating both modelling strategies depends on the inherent complex industrial oven conditions: If PLS2 prediction results are comparable to the standard PLS1 results; this will save a significant amount of the work for model updating and/or maintenance during uptime industrial operations. PLS2 requires only one calibration operation while PLS1 in this case would require 40 separate calibrations. If the single calibration based on PLS2 takes one hour, the PLS1 approach will take more than a week, so PLS2 is a much more efficient approach and a significant time saver in the industrial context, which needs to have a strong focus on minimal downtime.

2   Materials & methods

2.1   The hot air dryer

The hot-air-dryer has a length of 73 m and a width of 10 m. It is horizontally divided into 3 drying and 1 cooling zone, and is vertically stacked with three floors (Figure 1 and 2). While the zones are separated by hanging flaps, the floors are not separated, making the zones the operative unit in the drying process. It is up to the operators to make effective use of the vertical floors; there are significant advantages in planning the compound drying of several batches of boards (termed 'orders') intelligently. This requires a reliable PAT-based 'prediction of drying-end-results' facility, for which optimal monitoring of the compound drying will be critical.

The physical drying of the waterborne paint coating is based on the parameters air T, air speed, RH and drying time. The air is blown downwards onto the coating through tubes (Fig. 2).
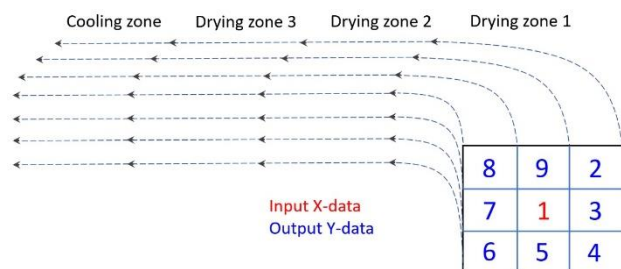
Figure 1. Simplified illustration of the nine drying trajectories for wood cladding. In the example indicated with red color, data from trajectory 1 was used as the input X data in calibration of the prediction model. The drying

conditions in all eight other trajectories were predicted based on the two alternative regression models PLS1 and PLS2.

As indicated with red color in Figure 1 the data from trajectory 1 was used as the input X data in calibration of the regression models. The drying condition described by the 40 output variables covering all the other eight trajectories were predicted based on the two alternative regression-modelling strategies PLS1 and PLS2, and compared see section 4.2.

Figure 2. The dryer in the coating plant, showing the upper, middle and lower floors in one drying zone. Note the air ducts close to the ceiling, which supply the drying air to all levels.

Partial Least Squares 1, PLS1 (single output) and Partial Least Squares 2, PLS2 (multiple y-outputs) were used to calibrate regression models for the drying conditions.

2.2   Sensors on the move

An innovative element in acquiring simultaneous data with which to perform a first model of the complex drying oven conditions was to send a recording '*sensor box*' through the entire drying cycle. The box includes a temperature and relative humidity (RH) sensor (S-THB-M002 Sensor, Onset Computer Corporation, Bourne, MA, USA) and three hot wire anemometers (T-DCI-F350-W5B3, Onset Computer Corporation) measuring the air flow horizontally in transport direction, horizontally perpendicular to the transport direction and vertically to the transport direction. All sensors have a maximum update rate of 400 ms. The measurements were recorded with a sampling rate of 1 Hz using a multi-channel data logger (Hobo H22, Onset Computer Corporation).
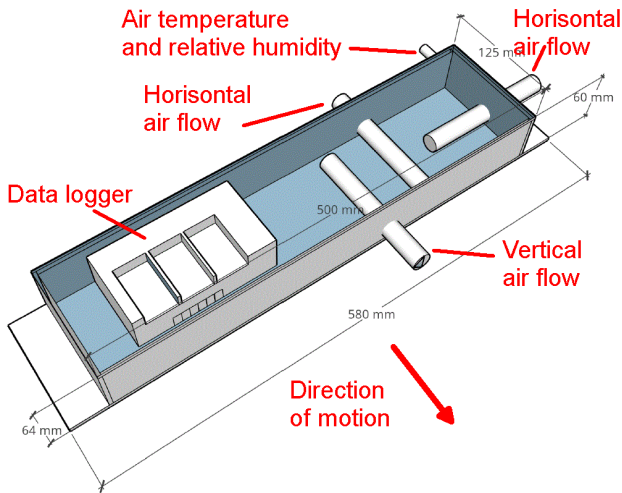
Figure 3. The moving "sensor box" equipped with *temperature/RH sensors and three hot-wire anemometers, measuring* along – and across the contemporary box velocity vector plus vertical air velocity.

2.3 Multivariate analysis and regression modelling

PLS1 regression is well described in the chemometric literature (Marten and Næs, 1989; Esbensen and Swarbrick, 2018) and numerous other background references herein. The PLS1 model is designed to maximize the effective correlations between the X-space (X-variables: PAT sensor recordings) and a singular y-variable. Here one also finds a description of the superior PLS modelling vs. prediction features in relationship to PCR (Principal Component Regression) and MLR (Multiple Linear Regression). The PLS1 approach has been the overwhelmingly most popular approach within process chemometrics because of the optimal modelling ability when dealing with one y-variable only.

Originally, in the early history of chemometrics, it was considered an advantage to be able to employ a PLS-regression modelling facility also in the case of a full multivariate Y-space (a multivariate set of y-variables), in cases where such a setup was required. The PLS2 modeling/prediction/validation situation is quite similar to PLS1, only the [X-space]-[Y-space] correlation modelling makes use of PLS components in both variable spaces simultaneously. In this fashion, the individual y_i variables are modelled based on the same foundational X-Y model. This only succeeds in the case in which all influential y-variables are strongly correlated – otherwise it is the overwhelming experience that modelling the individual y_i variables results in optimal prediction behavior – which is the reason for the dominant position of PLS1.

There only exist a few successful PLS2 application cases in the literature, e.g. Esbensen et al. (2001) who interconnected an acoustic chemometric approach using

'clamp-on' acoustic sensors (X) with laser velocimetry data in a radial pipeline profile. Their final feasibility demonstration successfully employed PLS2 inter-calibration of the acoustic monitoring data (X) with the laser velocimetric reference profiles (Y) using 39 y-variables.

A higher-level treatment of PLS2 method, including two practical application examples, can be found in (Stoccero et al., 2019).

With PLS1, the Y-data consist of a singular vector holding the 'reference values' used in calibration of the model. In PLS2, there are multiple y-output variables, so Y is a matrix in this case. Thus with PLS two alternative models are available, which is different from e.g. Linear regression and principal component regression (PCR) where only X is modelled; here no equivalent to PLS exists. The model for X and Y can be seen in equation (1) and (2) below.
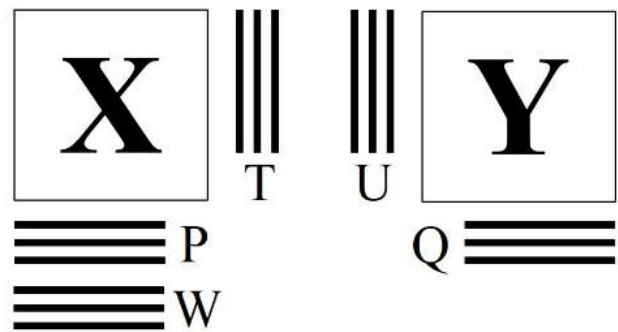
Figure 4. PLS data matrices (X and Y) and the resulting scores (T and U), loadings (P and Q) and Loading weights (W).

$$X = T \cdot P^T + E \qquad (1)$$
$$Y = U \cdot Q^T + F \qquad (2)$$

In the PLS approach the covariance between t and u are maximized which connects the two models.

PLS1 algorithm
Let $X_f = X$ and $y_f = y$ For f = 1, 2,…, A perform steps 2 to 6
1. $w_f = X^T y_f / \|X^T y_f\|$ (normalized to length 1)
2. $t_f = X \ w_f$
3. $q_f = t^T y_f (t_f^T t_f)^{-1}$
4. $p_f = X_f^T t_f (t_f^T t_f)^{-1}$
5. $b = u^T t (t^T t)^{-1}$
6. $X_{f+1} = X_f - t_f p_f^T$ and $y_{f+1} = y_f - b_f t_f^T$
7. $f{=}f{+}1$

Repeat 1 through 7 until f = A (the optimal number of components found by *validation*)

The compact regression coefficients for the Y variable can then be found like this:

$$\hat{b} = W(W^T X^T X W)^{-1} W^T X^T y \qquad (3)$$

Predictions based on new X-data can then be calculated:

$$\hat{y} = X\hat{b} + b_0 \qquad (4)$$

The PLS2 algorithm is similar but has a few more steps:

PLS2 algorithm
Let $X_f = X$ and $Y_f = Y$ For f = 1, 2,…, A perform steps 2 to 6
1. As the initial u_f vector, use any column in Y
2. $w_f = X^T u_f / |X^T u_f|$ (normalized to length 1)
3. $t_f = X w_f$
4. $q_f = Y^T t_f / |Y^T t_f|$ (normalized to length 1)
5. $u_f = Y q_f$
6. $p_f = X^T t (t^T t)^{-1}$
7. $b = u^T t (t^T t)^{-1}$
8. $X_{f+1} = X_f - t_f p_f^T$ and $Y_{f+1} = Y_f - b t_f q_f^T$
9. $f = f+1$

Repeat 1 through 9 until f=A (the optimal number of components found by validation)
The compact regression coefficients for the Y variables can again be found like this:

$$\hat{B} = W(W^T X^T X W)^{-1} W^T X^T Y \qquad (5)$$

Predictions based on new X-data can then be calculated:

$$\hat{Y} = X\hat{B} + B_0 \qquad (6)$$

In evaluating the regression model, the root mean squared error of prediction RMSEP, offset, slope and correlation coefficient are commonly used as model quality indices. Besides these, visual evaluation of the relevant score plots, loading weights plots, explained variance plots also provide useful information for calibrating and development of the prediction model (Marten and Næs, 1989; Esbensen & Swarbrick, 2018).

The root mean squared error of prediction is calculated as:

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_{i,\text{predicted}} - y_{i,\text{reference}})^2}{n}} \qquad (7)$$

3 Experimental

3.1 Air dryer experiments

The sensor box was mounted on a cladding board of 5518 mm length in three positions (left, middle, right) to record the climate across the width of the dryer. The position on both sides was 168 mm from the board end (Figure 5), and the box center was positioned 2759 mm from both ends. This was done for each of the three floors, giving nine runs (Figure 1 and Figure 2). After each run, the data were downloaded and deleted from the logger. The run-through time in each of the three drying zones was also recorded to assign the measurements to the zones. No other boards were in the dryer during the experiment.
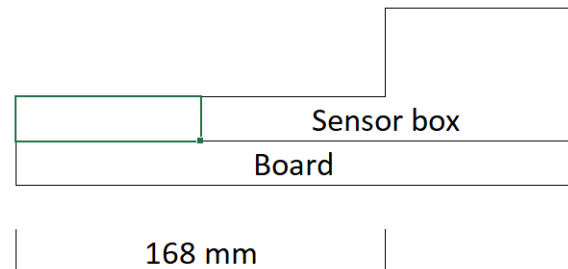
Figure 5. Positiong of the board and the box on the right side towards the gangway, trajectory 2 in Figure 2.

4 Results &discussion

4.1 PCA

The sensor recordings are represented as variables in the PCA X matrix, with the successive recording time points serve as objects. To get an overview of the full data set and how samples and variables relate to each other; PCA was used as the initial analysis method. All variables from all nine trajectories were included in this analysis.

The PCA score plot in figure 6 shows a marked depiction of the score value development with the three different drying zones (blue, red and green) clearly
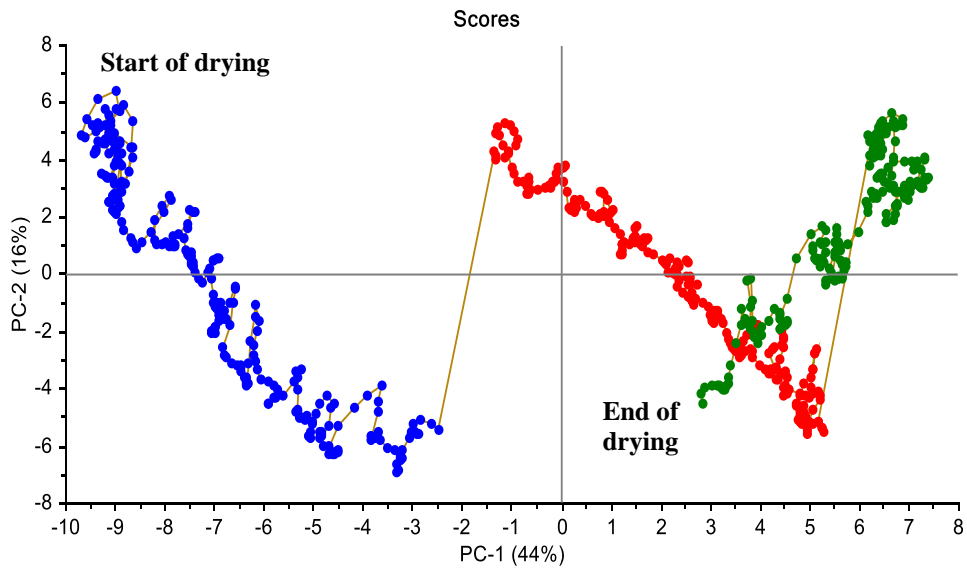
Figure 7. PCA – score plot of data from all nine drying trajectories as revealed by a full run-though in the drying oven of the "sensor box". Blue=drying zone 1, Red=drying zone 2 and Green=drying zone 3.
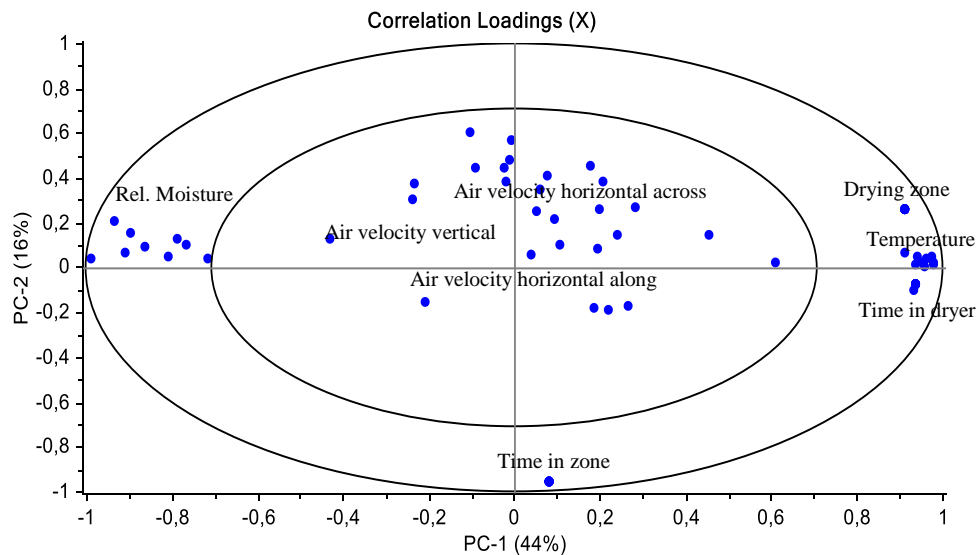
Figure 6. Correlation loadings corresponding to the score-plot in figure 6 (above) showing how the 8 variables from 9 drying trajectories relate to each other along PC1 and 2.

separated. No outliers can be seen in the data so all the samples and variables will be included in the next data modelling steps.

In the corresponding PCA loading plot, Fig. 7, all variables related to air velocity (vertical, horizontal along and horizontal across) are located close to the origin meaning their variance and covariance are less influential than the other variables occupying positions near the perimeter of this plot. These latter are the variables most influential on the model (variables that display a strong systematic inter-correlation), while the variables close to the origin contain a significant amount of irregular structure (noise).

4.2 PLS1 and PLS2 modelling results

Each variable in trajectory 2-9 was selected as an independent y-variable and was modelled by PLS1, which resulted in 40 separate PLS1 models. PLS2 was used to calibrate all the variables at once, which resulted in one PLS2 model also with 40 outputs.

Selection of training vs. validation data sets: The training data set was every second sample and the validation set was the other half of the data from the sensor box. Each PLS1 and PLS2 model was validated by this validation strategy (Esbensen and Swarbrick 2018). Results from the PLS1 and PLS2 models are reported and compared in Table 1 below.

Figure 8a shows the score-plot of the first two principal components (t1 vs t2). As can be seen it is possible to completely separate the samples from the three drying zones. No outliers were observed in the score plots t1-t2 nor in the score plots for the higher components.

The loading weights plot shows in figure 8b shows that all variables contribute to the model for the first three components. Although the variable "airflow horizontal across" shows relatively low loading values the prediction performance of the model does not improve if this variable is deleted.

Overfitting of regression models resulting from using too many components in the model is always a risk (Esbensen and Swarbrick 2018). As shown in the residual validation variance plot in figure 8c it is not immediately obvious how many components is optimal since the curve is decreasing for all the eight possible components. Since the model is validated based on independent data (test set validation), and there are 40 output variables to "satisfy" at once, we decided to use three components in the final prediction model, without loss of generality.

Figure 9 shows the predicted and reference data for a randomly selected trajectory, number 4. Figure 9a-c shows that the predicted air velocities can outline the main drying trends in the data well, but there is also a significant amount of irregular scatter along these trends, which we believe is due to the less influential air velocity data from the interior, very complex between-track and between-floors crosswinds set up by the ducted drying air supplied from the air ducts along the ceiling of the oven only.

We find these first foray predictions quite acceptable and maybe even better that the reference measurements recorded with the sensor box which contains a significant amount of random noise in themselves.

The plots of the predicted temperature and relative moisture content in figure 9d-e shows that the predictions follow the reference data very closely, which is promising for the eventual decision of implementing this modelling strategy in the industrial production.

Figure 8. PLS2 model, a) Score plot t1-t2-t3, b) Loading weights for component 1, 2 and 3, c) Residual validation variance.
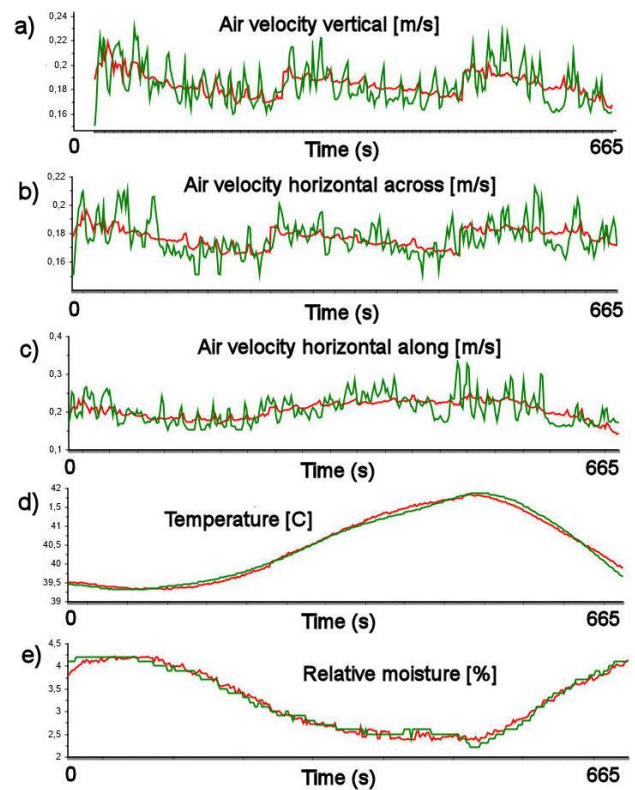
Figure 9. Predicted drying condition in trajectory 4 a) air velocity vertical b) air velocity horizontal across c) air velocity horizontal along d) Temperature e) Relative moisture content

A comparison of PLS 1 and PLS2 prediction results is shown in table 1. In all trajectories, the RMSEP values are similar for the two alternative modelling methods.

Table 1. PLS1 and PLS2 prediction comparison

Comparison of PLS1 and PLS2 model predictions			
Trajectory	Name of predicted variable	PLS1 RMSEP	PLS2 RMSEP
2	Air velocity vertical [mm/s]	10,34	10,38
2	Air velocity horizontal across [mm/s]	9,16	9,17
2	Air velocity horizontal along [mm/s]	34,60	34,63
2	Temperature [C]	0,04	0,04
2	Relative moisture [%]	0,15	0,15
3	Air velocity vertical [mm/s]	9,94	9,95
3	Air velocity horizontal across [mm/s]	7,93	7,96
3	Air velocity horizontal along [mm/s]	21,23	21,27
3	Temperature [C]	0,04	0,04
3	Relative moisture [%]	0,096	0,097
4	Air velocity vertical [mm/s]	12,28	12,33
4	Air velocity horizontal across [mm/s]	10,15	10,16
4	Air velocity horizontal along [mm/s]	28,59	28,62
4	Temperature [C]	0,085	0,085
4	Relative moisture [%]	0,09	0,09
5	Air velocity vertical [mm/s]	12,73	12,99
5	Air velocity horizontal across [mm/s]	8,95	9,06
5	Air velocity horizontal along [mm/s]	24,13	24,23
5	Temperature [C]	0,03	0,03
5	Relative moisture [%]	0,08	0,08
6	Air velocity vertical [mm/s]	13,04	13,19
6	Air velocity horizontal across [mm/s]	10,36	10,56
6	Air velocity horizontal along [mm/s]	17,98	18,15
6	Temperature [C]	0,06	0,06
6	Relative moisture [%]	0,09	0,09
7	Air velocity vertical [mm/s]	11,18	11,21
7	Air velocity horizontal across [mm/s]	10,73	10,91
7	Air velocity horizontal along [mm/s]	22,98	23,05
7	Temperature [C]	0,08	0,08
7	Relative moisture [%]	0,09	0,09
8	Air velocity vertical [mm/s]	11,24	11,56
8	Air velocity horizontal across [mm/s]	16,85	17,18
8	Air velocity horizontal along [mm/s]	17,12	17,33
8	Temperature [C]	0,06	0,06
8	Relative moisture [%]	0,20	0,20
9	Air velocity vertical [mm/s]	10,12	10,17
9	Air velocity horizontal across [mm/s]	7,14	7,62
9	Air velocity horizontal along [mm/s]	30,29	30,27
9	Temperature [C]	0,06	0,06
9	Relative moisture [%]	0,34	0,34

These results show it is possible to predict drying conditions in the dryer at GB with a satisfactory first foray quality. There were no significant differences for the diagnostic RMSEP values between PLS2 and PLS1 models. This indicates the possibility of establishing an effective maintenance schedule for such *in-production* models.

5 Conclusion

A feasibility study was carried out to assess the possibility of developing prediction models for monitoring drying conditions of wood coatings in one of Europe's largest and most modern dryers for exterior wood cladding. It was fully possible to perform detailed multivariate data models of the complex drying conditions and their interrelations using both Principal Component Analysis (PCA) and PLS-regression in both its PLS1 and PLS2 manifestations.

Since there were no significant differences for the diagnostic prediction performance RMSEP quality index between PLS2 and PLS1 models, PLS2 can be used to model the dryer at Gausdal Bruvoll, which indicates a positive prospect for very efficient data model updating and maintenance routine for future *in-production* PLS2 models. However, there remains a significant amount of industrial calibration work before a final evaluation can be performed.

References

K. A. Bakeev, *Process Analytical Technology, 2nd edition. Wiley*, Chichester, United Kingdom, 2010, ISBN: 978-0-470-72207-7, 2010. doi:10.1002/9780470689592

J. Beetsma. Alkyd Paints: From the ease of organic solvents to the difficulties of water. Proceedings of *XXIIth FATIPEC Conference*, Ease of Organic Solvents to the Difficulties of Water, 2: 157-164, 1995.

P. T. Elliot and J.E. Glass, *Applied Polymer Science: 21st Century, Elsevier Science*, 563-588, 2000. doi:10.1016/B978-0-08-043417-9.X5000-4

K. H. Esbensen, M. Waskaas, I. H. Matveyev, K. E. Wolden, J. G. Lode, T. T. Lied, and M. Halstensen. Feasibility of EMF-induced pipewall friction reduction by PLS2 inter-calibration of acoustic chemometrics and reference laser velocimetry. *J. Chemometrics*, 15: 241-263, 2001. doi: 10.1002/cem.684

K. H. Esbensen and B. Swarbrick, *Multivariate Data Analysis – An introduction to Multivariate Analysis, Process Analytical Technology and Quality by design*, 6th Ed., *CAMO Publishing*, 2018. ISBN 078-82-691104-0-1

U. Hundhausen, D. Kraniotis, S. Charisi, and M-L. Sortland. Quality control in coating lines for exterior cladding. From proceedings of *PRA's 10th International Woodcoatings Congress*, Amsterdam, The Netherlands, 2016. ISBN: 9783866302815

U. Hundhausen and M. Slabohm. In-line quality control of the film thickness in industrial coating processes for exterior cladding, Proceedings from the *International Research Group on Wood Preservation, Johannesburg, South Africa*, 18-40, 2018.

U. Hundhausen, M. Slabohm, and P. Meinlschmidt. Industrial coating of wood cladding: Inline control of board temperature, film thickness, and microfoam. Proceedings from *PRA's 11th International Woodcoatings Congress*, Amsterdam, The Netherlands, 23-24 October 2018. ISBN: 9783748600039

J. V. Koleske (Editor). *Paint and coating testing manual:15th Edition of the Gardner-Sward Handbook.* West Conshohocken, PA, USA, *ASTM International*, 2012. ISBN: 978-0-8031-7017-9. doi: 10.1520/MNL17-2ND-EB

H. Martens and T. Næs. *Multivariate Calibration*, *Wiley*, Chichester, UK, 1989 reprint 1994. ISBN-10: 0471930474, ISBN-13: 978-0471930471

J. G. Nienhuis, Review on drying and curing techniques of coatings. In proceedings from Cost E18 Final congress, Paris, 2004, 1-7, www.virtual.vtt.fi/virtual/proj6/coste18/nienhuisdryingpaper.pdf.

M. Stocchero, E. Locci, E. d'Aloja, M. Nioi, E. Baraldi, and G. Giordano. PLS2 in Metabolomics. Metabolites, 9(3):51, 2019. doi:10.3390/metabo9030051

A. Van Trent. *Turbidity Study of the Process of Film Formation of Polymer Particles in Drying Thin Film of Acrylic Latices.* PhD Thesis, University of Delft, The Netherlands, 1992.

J. W. Vanderhoff, E. B. Bradford, and W. K. Carrington. The transport of water through latex films. *Journal of Polymer Science*, *Polymer Symposia,* 41: 155-174, 1973. doi: 10.1002/polc.5070410116

Intelligent methodologies in recursive data-based modelling

Esko K. Juuso

Control Engineering, Environmental and Chemical Engineering, Faculty of Technology, University of Oulu, Finland
`esko.juuso@oulu.fi`

Abstract

Intelligent methodologies are beneficial in developing understandable multimodel simulation solutions. Nonlinear scaling extends these applications by facilitating compact nonlinear approaches already at the basic level. Composite local models can continue using linear methodologies for various case-based models. The flexible handling of the new structures and recursive tuning are the keys in adapting the systems in varying operating conditions. The recursive tuning of the scaling functions has two levels: smooth adaptation and strong shape changes. Fuzzy set systems further extend application areas of the models by combining composite local models in a flexible way. The extensions of the data-based methodologies are suitable for developing these adaptive applications via the following steps: variable analysis, linear models and intelligent extensions. Evolutionary computation is used in the tuning of the resulting complex models both the scaling and interactions. Complex problems are solved level by level to keep the domain expertise as an essential part.

Keywords: nonlinear systems, intelligent methods, recursive tuning, composite local models, linguistic equations

1 Introduction

In parametric models, the output is calculated as a linear combination of past inputs and past outputs. The output at time t can depend on many signal values y from previous time instants. The signals should be chosen according to the appropriate time delays. Fuzzy and neural models are based on the same structures and input-output models. Another solution for the operation changes with time is to use time-dependent proportional hazard models, where exponential functions are used to activate new hazards.

Linear methodologies extended with principal components (Jolliffe, 2002; Gerlach et al., 1979) and semi-physical models (Ljung, 1999) provide a feasible solution for many applications. Nonlinearities have been handled commonly with interaction and quadratic terms (Box and Wilson, 1951). Linear parameter varying (LPV) extend these solutions to decomposed systems (Hjartarson et al., 2015; Theis et al., 2018).

Artificial neural networks (ANNs) could be extended to highly complex architectures for handling complicated interactions within different sources of varying data (Schmidhuber, 2015). Fuzzy set systems can handle knowledge-based information (Zadeh, 1965; Takagi and Sugeno, 1985; Driankov et al., 1993; Dubois et al., 1999). Nonlinear dynamic models can be built by using fuzzy set systems (Babuška and Verbruggen, 2003) and low complexity neural networks (Sahoo et al., 2013). In hybrid dynamic models, local models need to be combined in complex systems (Sontag, 1981; Ljung, 2008; Jardine et al., 2006).

Linguistic equation (LE) models use the static mapping and NARX/Nonlinear AutoRegressive with eXogenous structures input models in the same way as fuzzy set systems and neural networks. The main difference is that the input and output variables are processed by a nonlinear scaling method, which originates from the membership functions used in fuzzy systems. (Juuso and Leiviskä, 1992; Juuso, 2004a) Constraints handling (Juuso, 2009a) and data-based analysis (Juuso and Lahdelma, 2010), improve possibilities to update the scaling functions recursively (Juuso, 2011). Different fuzzy approaches can be efficiently combined with LE models where the interactions between the scaled variables are linear (Juuso, 2014).

The LE approach integrates the knowledge-based systems, neural networks and evolutionary computation in the computational intelligence. Fuzzy set systems extend to parallel and contradictory phenomena in finding the operating areas, neural networks include more complex interactions, and evolutionary computation provide flexible optimization tools. The compact LE structures have many benefits: intelligent trend and deviation indices reveal changes and severity of them (Juuso, 2017). The nonlinear scaling is essential in dynamic models and prognostics, e.g. for predicting fatigue risk by using intelligent stress indices (Juuso and Ruusunen, 2013).

The recursive data analysis is based on two phases: first the parameters of the nonlinear scaling and then the model coefficients are analyzed to update the dynamic model. The aim is to keep the model equations unchanged as much as possible, i.e. interactions are not changing and the new situations are handled by changing the meanings of variables.

This paper focuses on the LE modelling approach based on nonlinear scaling (Section 2) enhanced with recursive solutions of updating the scaling and model interactions (Section 3). Composite local models and intelligent methodologies are combined in the multimodel LE simulation (Section 4). Possibilities of the new structures and recursive scaling are analyzed within a large set of previous applications in Section 5. Experiences and new pos-

sibilities are discussed in Section 6 and conclusions and future research are presented in Section 7.

2 Modelling methodologies

Nonlinear scaling and steady-state statistical modelling with linear methodologies are the basis of the LE modelling. The fuzzy meanings of the variables are embedded in the nonlinear scaling. Interactions are analyzed by linear models and dynamic models are based on additional structures.

2.1 Nonlinear scaling

The nonlinearities of the process are handled by the nonlinear scaling of the variables. The scaling functions are monotonously increasing functions $x_j = f(X_j)$ where x_j is the variable and X_j the corresponding scaled variable. The function $f()$ consist of two second order polynomials, one for the negative values of X_j and one for the positive values, respectively. The corresponding inverse functions $x_j = f^{-1}(X_j)$ based on square root functions are used for scaling to the range [-2, 2], denoted linguistification. In LE models, the results are scaled to the real values by using the function $f()$. (Juuso, 2004a)

The support area is defined by the minimum and maximum values of the variable, i.e. the support area is $[\min(x_j), \max(x_j)]$ for each variable $j, j = 1, \ldots, m$. The central tendency value, c_j, divides the support area into two parts, and the core area is defined by the central tendency values of the lower and the upper part, $(c_l)_j$ and $(c_h)_j$, correspondingly. This means that the core area of the variable j defined by $[(c_l)_j, (c_h)_j]$ is within the support area. The parameters of the functions are extracted from measurements by using generalised norms and moments (Juuso and Lahdelma, 2010).

2.2 Steady-state modelling

The steady-state simulation models are linear *multiple input, multiple output (MIMO)* models $\vec{y} = F(\vec{x})$, where the output vector $\vec{y} = (y_1, y_2, \ldots, y_n)$ is calculated by a linear function F from the input vector $\vec{x} = (x_1, x_2, \ldots, x_m)$. Statistical modelling in its basic form uses linear regression for solving coefficients for a linear function. Linear methodologies are suitable for large multivariable systems. Quadratic and interactive terms are not used here. Principal components compress the data by reducing the number of dimensions with a minor loss of information (Jolliffe, 2002). Partial least squares regression (PLS) is an extension of these ideas (Gerlach et al., 1979). Known semi-physical models of inputs are important in linear modelling, see (Ljung, 1999).

The directions of the interactions analyzed with these methodologies are aimed to be valid in a wider area than the scaling functions of the model variables. The LPV modelling further extends the feasible areas of the model parameters.

2.3 Dynamic modelling

Data-driven modelling with parametric models, also known as identification (Ljung, 1999), is the key methodology in the dynamic modelling. In NARX models, the input and output values are chosen according to appropriate system orders. The regressor vector consists of a finite number of past inputs and outputs. The number may become too high in nonlinear systems. The dynamic structures are reduced in dynamic models based on fuzzy set systems or neural networks. The nonlinear scaling further reduces the number of input and output signals needed for all these systems.

3 Recursive modelling

Recursive data analysis facilitates the adaptation of the functions to changing operating conditions: the parameters of the scaling functions are obtained by using the earlier analyzed norms, also the orders of the norms can be re-analyzed if needed. Machine learning can be used to decide if the recursive updates are needed.

3.1 Recursive nonlinear scaling

The parameters of the nonlinear scaling functions can be recursively updated by including new equal sized subblocks in the calculations, i.e. the norm for several samples can be obtained as the norm of the norms of the individual samples:

$$||{}^{K_s\tau}M_j^p||_p = \{\frac{1}{K_s}\sum_{i=1}^{K_s}[({}^{\tau}M_j^p)_i^{1/p}]^p\}^{1/p} = [\frac{1}{K_s}\sum_{i=1}^{K_s}[({}^{\tau}M_j^p)_i]^{1/p},$$

(1)

where K_s is the number of samples $\{x_j\}_{i=1}^N$. In automation and data collection systems, the sub-blocks are normally used for arithmetic mean ($p = 1$). The norm values can be recursively updated with (1), and a new search for the orders is done only if the values change considerably (Juuso, 2011). The number of samples can be increasing or fixed with some forgetting, and weighting of the individual samples can be used in the analysis. The nonlinear scaling need to be done by a set of functions when the nonlinearities are very strong (Juuso and Lahdelma, 2010) and operating condition specific.

The analysis is the same for any combination of the subblocks. The central tendency value is chosen by the point where the skewness changes from positive to negative, i.e. $\gamma_3^p = 0$. Then the data set is divided into two parts: a lower part and an upper part. The same analysis is done for these two data sets. The estimates of the corner points, $(c_l)_j$ and $(c_h)_j$, are the points where $\gamma_3^p = 0$ for the lower and upper data sets, respectively. Since the search of these points is performed by using the order of the moment, the resulting orders $(p_l)_j$, $(p_0)_j$ and $(p_h)_j$ are good estimates when additional data sets are used.

There is all the time an active set of scaling functions (Figure 1. The functions are tuned to appropriate operating areas by using the corresponding sets of data collected
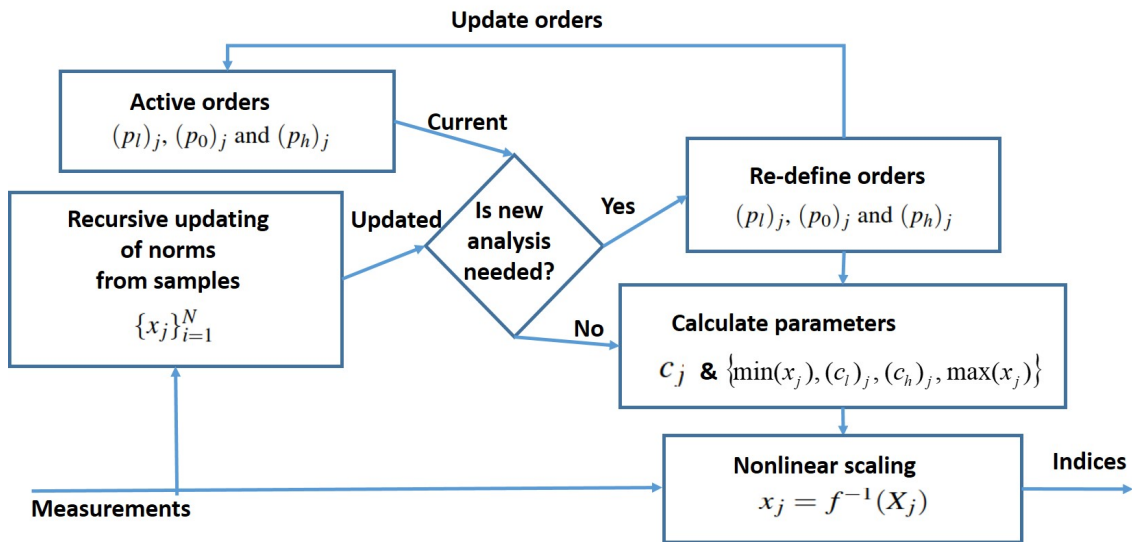
Figure 1. Recursive updating of scaling functions.

from the sub-blocks of the previous periods. The feasibility ranges should be wider for the working point variables. There are two levels of complexity for updating the scaling functions:

1. Individual scaling functions are smoothly adapted by using the existing norms (1) for the new data. This level requires only the values for the norms with the orders $(p_l)_j$, $(p_0)_j$ and $(p_h)_j$.

2. The shape of the functions is modified by recursively updating the norm orders. In this case, the orders of the norms are updated from the data of the sub-blocks.

In both methodologies, the monotonous increase of the functions are checked and corrected if needed.

The scaling functions of the working point variables are aimed to cover the whole modelling area, i.e. they are analyzed with the same methodology, but for the set of the data includes several operating conditions.

3.2 Interactions

The model parameters can be updated by re-running the regression after the variables, whose scaling functions have been recursively updated, have been re-scaled. This is done for specific operating areas by using appropriate data. The procedure is the same for all submodels and working point models in any operating area.

Different operating areas can be analyzed with clustering to find feasible areas for local models. Cluster analysis provides hundreds of algorithms for the data-driven analysis (Xu and Tian, 2015). Iterations are needed since the clustering algorithms work better with linear systems, i.e. after nonlinear scaling.

4 Multimodel LE simulation

Multimodel approaches divide the problem into smaller parts for developing separate models for subprocesses or different stages in the process operation interconnected with process streams. Additional properties are achieved because equations and delays are allowed to vary between different submodels. The working areas can be defined by a separate working point model. The submodels are developed using the case-based modelling approach.

4.1 Composite local models

The composite local models use linear approximations of the nonlinear system in different neighbourhoods. If the partitioning is based on working point variables, the partitioning can be used in weighting the local models. In linear parameter varying (LPV) models, the matrices of the state-space models depend on an exogenous variable measured during the operation (Hjartarson et al., 2015; Theis et al., 2018). Piecewise affine (PWA) systems extend the local linear models to a polyhedral partition where the models can be state-space or parametric models (Christophersen, 2007). The model switches between different modes as the state variable varies over the partition. A high number of local models brings an overfitting risk.

Working point variables and their interactions can form a working point index for LPV models which are useful in smoothing the operation in varying operating conditions. The number of specific PWA systems can be reduced by using fuzzy set systems.

The LE models are defined by the parameters of the scaling functions and the coefficients of the interaction models. The idea of the exogenous variables can be used for these parameters, which opens a set of new modelling approaches for the nonlinear parameter (NPV) varying models where the exogenous variables are extended with
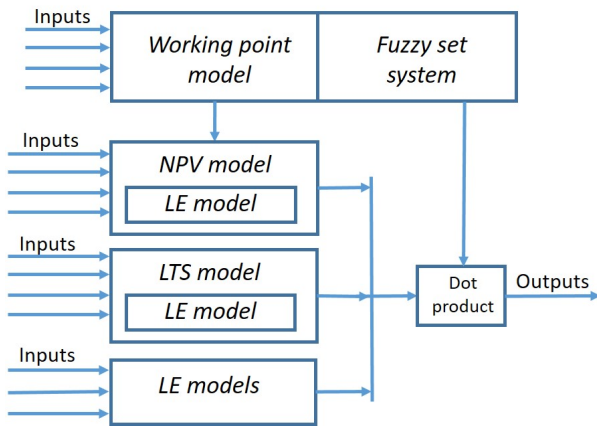
Figure 2. Multimodel LE system with a fuzzy decision module.

working point models (Figure 2).

4.2 Intelligent systems

The multimodel LE system consists of different types of submodels, a working point model and a fuzzy set system (Figure 2). The basic form includes a set of linguistic equations where each equation is specific for a certain phenomena which can be active in several operating conditions. The working point model calculates the degree of membership for each operating area which taken as a degree of membership for all the active equations of the case. Each model produces an alternative solution whose degree of membership $[0, 1]$ is obtained by the fuzzy set system. All alternative solutions are aggregated and the response is defuzzified. Since the nonlinear scaling is operating condition specific, each active equation of each active operating condition produces an alternative result.

For certain phases of the process, the submodels in Figure 2 can be composite local models which extend the operating area of the submodels. The NPV models include a nonlinear parameter handling based on the LE working point models, i.e. the coefficients of the equations depend on the operating conditions. As much as possible is done with composite models and fuzzy methodologies are used for uncertainty and decisions.

Fuzzy set systems expand composite local models to partially overlapping models. Fuzzy models combine local modelling approaches and facilitate gradual changes. Takagi-Sugeno (TS) fuzzy models are based on linear submodels. The smoothing problems around the submodel borders need special techniques, e.g. smoothing maximum, or by making the area overlap very strong as it is done in the ANFIS method (Jang, 1993) which is practice a smoothing algorithm.

In the multimodel LE systems, a fuzzy decision system is used for selecting and weighting suitable nonlinear submodels for each situation (Figure 2). The fuzzy system based on a working point model or a single working point index provides several alternatives with different degrees

of membership. In Linguistic Takagi-Sugeno fuzzy models (LTS), the fuzzy partition is defined with the same variables as the models and the nonlinearities are handled with the scaling functions and the interaction part with fuzzy set systems. The multimodel can be developed and tuned with the same methods as the normal TS models (Juuso, 2009b).

The fuzzy calculations are beneficial if there are parallel gradually activating phases or contradictory models. Decomposition is needed to extend the solutions to different subprocesses, process phases, phenomena and multiple operating conditions. In addition to spatial or logical blocks, the decomposed modelling can be based on different frequency ranges. The mixed systems may also include models based on the first principles.

In deep neural networks, an ensemble of redundant networks improves generalisation by averaging the process of creating multiple models and combining them to produce a desired output (Xiao et al., 2018). Naturally, this methodology could be used together with the nonlinear scaling. However, the fuzzy methodologies are preferred to keep the understanding of the subsystems strong.

4.3 Evolutionary computing

Evolutionary computing is widely used to tune intelligent systems which incorporate expert knowledge with data. Genetic algorithms (GA) are well suited for LE models based on nonlinear scaling and linear interactions. The scaling functions handle efficiently the parameter constraints of the monotonously increasing second order polynomials and the whole system is configured with a set of parameters. (Juuso, 2009a)

All the interactions within the multimodel LE system (Figure 2) are represented with compact linear equations whose coefficients can be included in the tuning based with genetic algorithms. The coefficients of the equations are re-analyzed for different operating conditions after updating the scaling functions.

The GA approach is flexible: an appropriate set of parameters is taken in the optimization and the performance is evaluated with the same methodology in all these levels. The recursive tuning of the scaling functions can be done with GAs without using the data-driven approach presented in Section 3.1.

5 Applications

Nonlinear scaling forms the basis for the LE modelling: an important benefit of the linear approach is that the models can be inverted, technically to any direction. The compact basic solution makes extensions to dynamic and case-based systems possible. Complex models for steady-state and dynamic systems can be built with the cascade and interactive structures.

Possibilities of the recursive tuning are analysed for the earlier applications presented in (Juuso, 2018). The topics of these applications are explained with more details in papers referred in this section. This paper focuses on how

Table 1. Steady-state LE model applications.

Case	Application area	Modelling	Recursive analysis
Electric furnace	DSS for process design	Nonlinear models transformed to LE models and Interactions	Raw materials
Lime kiln	Feedforward control	Fuel feed in changing capacity conditions	Fuel properties
Solar collector field	Control adaptation	Working point model: irradiation, temperature, difference, special cases with fuzzy set systems	Deterioration of process condition
Continuous cooking	Quality control	Quality forecasting	Raw material variations
Fatigue	Stress contributions	LE based stress-cycle curve: -2nd order scaling for the stress -logarithmic scaling for the cycles	Deterioration of process condition
Water treatment	Feedforward control	Turbidity for control Forecasting residual aluminium	Seasonal effects
Wastewater treatment	Diagnostics	Operating conditions	Incoming wastewater purity

Table 2. Dynamic LE model applications.

Case	Application area	Modelling	Recursive analysis
Gas furnace	Modelling	Development and tuning: training, validation, testing	Fluctuations
Solar collector field	Controller tuning	Time varying transport delay Cloudy periods	Deterioration of the condition
Fatigue	Forecasting fatigue risk	Cumulative sum of scaled stress contributions Rolling mill, LHD machines	Deterioration process condition
Water treatment	Controller tuning	Water quality indicator Water circulation Drinking water	Seasonal effects
Condition monitoring	Prognostics	Recursive tuning	Deterioration of the condition

the applications can be improved with the new structures and recursive tuning, see the column *Recursive analysis* in Tables 1, 2 and 3.

5.1 Steady-state LE models

Steady-state LE models are mainly used in adaptation and feedforward control (Table 1). In most cases, the models including only a single linear equation can be expanded in a straightforward way. The first LE model developed for designing submerged arc furnaces was an exception which used well known relations represented by five equations (Juuso and Leiviskä, 1992). Variations of raw material properties can be compared within additional levels. A steady-state LE model was developed from the process measurements in an early lime kiln control application where varying fuel properties require attention (Juuso et al., 1997). The working point model is an essential part

of the model-based LE control of a solar power plant (Juuso and Yebra, 2013). During the years, the field condition has deteriorated, which has required recursive updates.

For continuous cooking, a LE model was developed for predicting the Kappa number, which is widely used quality variable (Leiviskä et al., 2001). Raw material variation should be taken into account.

Stress-cycle (S-N) curves, also known as Wöhler curves, are represented by a compact linguistic equation (Juuso and Ruusunen, 2013). Machine deterioration changes the meanings of the stress levels.

In drinking water applications, models have been developed for forecasting and control (Tomperi et al., 2013). Seasonal effects are important in these applications. Operating conditions are detected in diagnosing the wastewater treatment process (Juuso and Laakso, 2013). The strong effect of the biomass performance requires recursive up-

Table 3. Decomposed LE model applications.

Case	Application area	Modelling	Recursive analysis
Lime kiln	Fuel quality	Controller tuning by using multiple models	Fluctuations in raw materials
	Adaptive control	Working point control	Varying operating conditions are essential
Solar collector field	Controller tuning for oil flow	Models for different operating conditions	Irradiation fluctuations
		Distributed parameter models	Varying situations at the collector field
Batch cooking	On-line forecasting	Three interactive models: alkali, lignin and dissolved solids	New structure for interactions and phases
Fluidised bed granulation	Forecasting	Three interactive models: temperature, humidity and granular size	New structure interactions and phases
Fed-batch fermentation	On-line forecasting	Submodels of three growth phases, each including three interactive models	New structure for interactions and phases
		Totally nine interactive models	Activation and decline of partially simultaneous processes
Wastewater treatment	Detection of operating conditions	Three submodels: load, treatment and settling	Biomass performance
		Trend analysis	

dates.

5.2 Dynamic LE models

The basic dynamic LE model is represented by a compact parametric model

$$Y(t) + a_1 Y(t-1) = b_1 U(t - n_k) + e(t) \qquad (2)$$

for the scaled variables Y and U. More complicated input-output NARX models containing are simplified by using the nonlinear scaling. The models can have an appropriate number of variables. The approach was first tested in a gas furnace data provided by (Box and Jenkins, 1970). The dynamic models of the solar plant are based on test campaigns, which cannot be planned in detail because of changing weather conditions (Juuso, 2003a). The irradiation changes and varying cloudy conditions have strong effects.

The basic dynamic flotation model is the core of the quality indicator in water treatment (Joensuu et al., 2005). A dynamic LE model has been used for the fatigue prediction in (Juuso and Ruusunen, 2013). In all these models, only one equation is needed. The applications are indirect measurements and controller tuning (Table 2). Drinking water applications focus on model-based control and forecasting (Tomperi et al., 2016). The multimodel structures (Juuso et al., 2009) and trend analysis (Tomperi et al., 2017) are important in the wastewater treatment.

The new structure and recursive tuning allow the handling of deterioration of condition, fluctuations and seasonal effects (Table 2). The recursive tuning was already used in the condition monitoring application. In most real applications, the changes are too fast to use recursive updates online. The offline analysis expands the libraries for predefined adaptation.

5.3 Decomposition in LE models

The multimodel LE system can include several submodels and complex interactions (Table 3). All basic submodels are represented by (2) extended to an appropriate number of variables. The model with a fuzzy decision module was first used for a lime kiln (Juuso, 1999) and then for a solar thermal power plant (Juuso, 2003a). The lime kiln model had six operating areas defined by the production level and the trend of the fuel feed (increasing, decreasing). The model of the collector field includes four operating areas: start-up, low, normal and high operation. Additional fuzzy models have been developed for special situations (Juuso et al., 2000). In these cases, the decision module use working point variables are provides specific degrees of membership for all the submodels. Recursive tuning is needed only for specific submodels in new situations.

Interactive dynamic models were needed in several cases: batch cooking (Juuso, 2003b), fluidised bed granulator (Mäki et al., 2004), industrial fed-batch fermenter (Saarela et al., 2003) and wastewater treatment (Juuso et al., 2009). Linguistic equations, neural networks and fuzzy modelling with several variants have been compared by using the process data obtained from the fed-batch fermenter. The LE models are sufficiently compact to be used in the complex system presented in Figure 2.

The phases and interactive submodels make the modelling complicated (Table 3). Additional requirements were recently detected for the fed-batch fermentation case, where the activation and decline of different phenomena need be taken into account (Juuso, 2019). The new structure of integrating composite submodels and intelligent systems (Figure 2) and recursive tuning provide improvements for the handling of interactions and phases as well as the fluctuations of fuel properties and performance of biomass in the treatment.

Phenomenological models can be integrated with these solutions through a data-driven model which provides the parameters needed in the models of phenomena.

5.4 Distributed parameter LE models

In the distributed parameter models, the solar collector field is divided into modules, where the dynamic LE models are applied in a distributed way (Juuso, 2004b). The same single equation model (2) with an appropriate number of variables is used in all modules. Element locations for partial differential equations (PDEs) are defined by the flow rate. In cloudy conditions, the heating effect can be strongly uneven.

The new structure and recursive tuning provide possibilities to handle local variations of the collector field which become important if the temperature increase is high (Table 3).

6 Discussion

The nonlinear scaling methodology is the key in the extensions of the linear methodologies. The scaling can be recursively updated and used in both steady-state and dynamic applications with additional structures. Several steady-state and dynamic models are combined with fuzzy set systems. Distributed parameter systems can use the same algorithms. In the applications discussed in Section 5, the scaling functions were developed before the current data-based analysis. New structures and recursive tuning provide additional possibilities for them.

The variable specific recursive analysis of the parameters of the scaling functions is feasible throughout the modelling. This is important in the smooth adaptation of the submodels and introducing models for new operating areas. The interactions are tried to keep unchanged but can be retuned when needed.

The multimodel structure facilitates deep learning extensions. All parameters of the multimodel LE systems, including any subsystem or model, can be updated with the same genetic tuning. Constraints are taken into account in the coding which means that penalties are not needed in the optimisation.

The recursive approach is in this paper presented as an offline analysis. The methodology is applicable for the online analysis as well. Statistical process control provides additional tools for detecting changes, anomalies and novelties. The existing scaling functions provide a basis for

assessing the quality of new data: outliers should be excluded, but the suspicious values may mean that the operating conditions are changing and the support area should be extended. The online analysis could be used in some subsystems in specific operating areas.

7 Conclusions and future research

The nonlinear scaling approach extends the application areas of linear methodologies to nonlinear modelling: the meanings of variables and interactions are analysed either sequentially or simultaneously. Local nonlinear models and recursive tuning reduce the number of the local models. Intelligent methodologies are essential parts of the system: the close connection to the fuzzy set systems provides a good basis for understandable models and evolutionary computation is used in the tuning of the resulting complex models. Composite local models and intelligent systems can be efficiently combined. The basic models are compact and additional properties, including dynamics, uncertainty and decomposition, are included if needed. The recursive analysis is a basis for future research in various applications discussed in this paper.

References

R. Babuška and H. Verbruggen. Neuro-fuzzy methods for nonlinear system identification. *Annual Reviews in Control*, 27 (1):73–85, 2003.

G. E. P. Box and G. M. Jenkins. *Time Series Analysis, Forecasting and Control*. Holden Day, San Francisco, 1970.

G. E. P. Box and K. B. Wilson. On the experimental attainment of optimum conditions. *Journal of the Royal Statistical Society. Series B*, 13(1):1–45, 1951.

F. J. Christophersen. Piecewise Affine Systems. In M. M. Gupta and T. Yamakawa, editors, *Optimal Control of Constrained Piecewise Affine Systems, Lecture Notes in Control and Information Sciences*, volume 359, pages 39–42. Springer, Berlin, Heidelberg, 2007. doi:10.1007/978-3-540-72701-9_4.

D. Driankov, H. Hellendoorn, and M. Reinfrank. *An Introduction to Fuzzy Control*. Springer, Berlin, Germany, 1993.

D. Dubois, H. Prade, and L. Ughetto. Fuzzy logic, control engineering and artificial intelligence. In H. B. Verbruggen, H.-J. Zimmermann, and R. Babuska, editors, *Fuzzy Algorithms for Control, International Series in Intelligent Technologies*, pages 17–57. Kluwer, Boston, 1999.

R. W. Gerlach, B. R. Kowalski, and H. O. A. Wold. Partial least squares modelling with latent variables. *Anal. Chim. Acta*, 112(4):417–421, 1979.

A. Hjartarson, P. Seiler, and A. Packard. LPVTools: A toolbox for modeling, analysis, and synthesis of parameter varying control systems. *IFAC-PapersOnLine*, 48(26):139–145, 2015. URL www.scopus.com.

J.-S. R. Jang. ANFIS: Adaptive-Network-based Fuzzy Inference Systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(3):665–685, 1993.

A. K. S. Jardine, D. Lin, and D. Banjevic. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7):1483–1510, 2006.

I. Joensuu, M. Piironen, and E. Juuso. Dynamic simulator for dosing of water treatment chemicals. *Computer-aided chemical engineering, European Symposium on Computer Aided Process Engineering-15 (Escape-15), Barcelona, Spain, May 29 - June 1, 2005*, 20A:301–306, 2005.

I. T. Jolliffe. *Principal Component Analysis*. Springer, New York, 2 edition, 2002. 487 pp.

E. Juuso and I. Laakso. Hybrid LE systems for simulation of an activated sludge process. In *Proceedings - 8th EUROSIM Congress on Modelling and Simulation, EUROSIM 2013, 10-13 September, 2013, Cardiff, UK*, pages 124–129, 2013. doi:10.1109/EUROSIM.2013.32.

E. Juuso and S. Lahdelma. Intelligent scaling of features in fault diagnosis. In *7th International Conference on Condition Monitoring and Machinery Failure Prevention Technologies, CM 2010 - MFPT 2010, 22-24 June 2010, Stratford-upon-Avon, UK*, volume 2, pages 1358–1372, 2010. URL www.scopus.com.

E. Juuso and M. Ruusunen. Fatigue prediction with intelligent stress indices based on torque measurements in a rolling mill. In *10th International Conference on Condition Monitoring and Machinery Failure Prevention Technologies, CM 2013 - MFPT 2013, 18-20 June 2013, Krakow, Poland*, volume 1, pages 460–471, 2013. URL www.scopus.com.

E. Juuso, T. Latvala, and I. Laakso. Intelligent analysers and dynamic simulation in a biological water treatment process. In I. Troch and F. Breitenecker, editors, *6th Vienna Conference on Mathematical Modelling - MATHMOD 2009, February 11-13, 2009, Argesim Report no. 35*, pages 999–1007. Argesim, 2009. ISBN 978-3-901608-35-3.

E. K. Juuso. Intelligent dynamic simulation of a lime kiln with linguistic equations. In H. Szczerbicka, editor, *ESM'99: Modelling and Simulation: A tool for the Next Millenium, 13th European Simulation Multiconference, Warsaw, Poland, June 1–4, 1999*, pages 395–00, Delft, The Netherlands, 1999. SCS.

E. K. Juuso. Intelligent dynamic simulation of a solar collector field. In A. Verbraeck and V. Hlupic, editors, *Simulation in Industry, 15th European Simulation Symposium ESS 2003*, pages 443–449. SCS, Gruner Druck, Erlangen, Germany, 2003a.

E. K. Juuso. Intelligent dynamic simulation of batch cooking. In *44th Scandinavian Conference on Simulation and Modeling SIMS 2003, September 18 -19, 2003*, pages 163–169. Malardalen University, Västeras, Sweden, 2003b.

E. K. Juuso. Integration of intelligent systems in development of smart adaptive systems. *International Journal of Approximate Reasoning*, 35(3):307–337, 2004a. doi:10.1016/j.ijar.2003.08.008.

E. K. Juuso. Dynamic simulation of a solar collector field with intelligent distributed parameter models. In B. Elmegaard, Jon Sporring, Kenny Erleben, and Kim Soerensen, editors, *Proceedings of SIMS 2004 - the 45th Scandinavian Conference on Simulation and Modelling, September 23-24, Copenhagen, Denmark*, pages 141–153. DTU, Lungby, Denmark, 2004b.

E. K. Juuso. Tuning of large-scale linguistic equation (LE) models with genetic algorithms. In M. Kolehmainen, editor, *Revised selected papers of the International Conference on Adaptive and Natural Computing Algorithms - ICANNGA 2009, Kuopio, Finland, Lecture Notes in Computer Science*, volume LNCS 5495, pages 161–170. Springer-Verlag, Heidelberg, 2009a. doi:10.1007/978-3-642-04921-7_17.

E. K. Juuso. Development of multiple linguistic equation models with Takagi-Sugeno type fuzzy models. In J. P. Carvalho, D. Dubois, U. Kaymak, and J. C. M. Sousa, editors, *Proceedings of 2009 IFSA World Congress / 2009 EUSFLAT Conference, July 20-24, 2009, Lisboa, Portugal*, pages 1779–1784. IFSA-EUSFLAT, 2009b. http://www.eusflat.org/publications_proceedings_IFSA-EUSFLAT_2009.php.

E. K. Juuso. Recursive tuning of intelligent controllers of solar collector fields in changing operating conditions. In S. Bittani, A. Cenedese, and S. Zampieri, editors, *Proceedings of the 18th World Congress The International Federation of Automatic Control, Milano (Italy) August 28 - September 2, 2011*, pages 12282–12288. IFAC, 2011. doi:10.3182/20110828-6-IT-1002.03621.

E. K. Juuso. Intelligent methods in modelling and simulation of complex systems. *Simulation Notes Europe SNE*, 24(1):1–10, 2014. doi:10.11128/sne.24.on.102221.

E. K. Juuso. Intelligent performance analysis with a natural language interface. *Management Systems in Production Engineering*, 25(3):168–175, 2017. doi:10.1515/mspe-2017-0025.

E. K. Juuso. Intelligent multimodel simulation in decomposed systems. In L. E. Øi, T. Komulainen, R. T. Bye, and L. O. Nord, editors, *Proceedings of The 59th Conference on Simulation and Modelling (SIMS 59), 26-28 September 2018, Oslo Metropolitan University, Norway*, number 153 in Linköping Electronic Conference Proceedings, pages 308–315. Linköping University Electronic Press, 2018. doi:10.3384/ecp18153308.

E. K. Juuso. Intelligent dynamic simulation of fed-batch fermentation processes. In E. Dahlquist, E. Juuso, B. Lie, and L. Eriksson, editors, *Proceedings of The 60th Conference on Simulation and Modelling (SIMS 60), 13-16, 2019, Västerås, Sweden*, number 170 in Linköping Electronic Conference Proceedings, pages 132–138. Linköping University Electronic Press, Linköpings universitet, 2019. doi:10.3384/ecp20170132.

E. K. Juuso and K. Leiviskä. Adaptive expert systems for metallurgical processes. In S.-L. Jämsä-Jounela and A. J. Niemi, editors, *Expert Systems in Mineral and Metal Processing, IFAC Workshop, Espoo, Finland, August 26-28, 1991, IFAC Workshop Series, 1992, Number 2*, pages 119–124, Oxford, UK, 1992. Pergamon.

E. K. Juuso and L. J. Yebra. Model-based intelligent control of a solar energy collector field. In *Proceedings - 8th EUROSIM Congress on Modelling and Simulation, EUROSIM 2013, 10-13 September, 2013, Cardiff, UK*, pages 513–518, 2013. doi:10.1109/EUROSIM.2013.92. URL www.scopus.com.

E. K Juuso, T. Ahola, and K. Leiviskä. Fuzzy modelling of a rotary lime kiln using linguistic equations and neuro-fuzzy methods. In L. Foulloy, editor, *Proceedings of the 3rd IFAC Symposium on Intelligent Components and Instruments for Control Applications - SICICA'97, Annecy, France, June 9-11, 1997*, pages 579–584. Pergamon, 1997. ISBN 9780080426082.

E. K. Juuso, D. Schauten, T. Slawinski, and H. Kiendl. Combination of linguistic equations and the fuzzy-ROSA method in dynamic simulation of a solar collector field. In L. Yliniemi and E. Juuso, editors, *Proceedings of TOOLMET 2000 Symposium - Tool Environments and Development Methods for Intelligent Systems, Oulu, April 13-14, 2000*, pages 63–77, Oulu, 2000. Oulun yliopistopaino.

K. Leiviskä, E. Juuso, and A. Isokangas. Intelligent modelling of continuous pulp cooking. In K. Leiviskä, editor, *Industrial Applications of Soft Computing*, Studies in Fuzziness and Soft Computing, pages 147–158. Springer, Heidelberg, 2001. doi:10.1007/978-3-7908-1822-2_10.

L. Ljung. *System Identification - Theory for the User*. Prentice Hall, Upper Saddle River, N.J., 2nd edition, 1999.

L. Ljung. Perspectives on system identification. In M. J. Chung and P. Misra, editors, *Plenary papers, milestone reports & selected survey papers, 17th IFAC World Congress, Seoul, Korea, July 6-11, 2008*, pages 47–59. IFAC, 2008. http://www.ifac-papersonline.net/.

T. Mäki, E. Juuso, and K. Leiviskä. Fuzzy modelling and dynamic simulation of a fluidised bed granulator. In K. Leiviskä, editor, *Proceedings of AFNC'04 - the 2nd IFAC Workshop on Advanced Fuzzy/Neural, September 16-17, 2004, Oulu, Finland*, pages 133–138. Finnish Automation Society, Helsinki, 2004.

U. Saarela, K. Leiviskä, E. Juuso, and A. Kosola. Modelling of a fed-batch enzyme fermentation process. In *IFAC International Conference on Intelligent Control Systems and Signal Processing. Faro, Portugal, April 8-11, 2003*. IFAC, 2003.

H.K. Sahoo, P.K. Dash, and N.P. Rath. Narx model based nonlinear dynamic system identification using low complexity neural networks and robust H∞ filter. *Applied Soft Computing*, 13(7):3324 – 3334, 2013. ISSN 1568-4946. doi:10.1016/j.asoc.2013.02.007.

J. Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61 (Supplement C):85 – 117, 2015. ISSN 0893-6080. doi:https://doi.org/10.1016/j.neunet.2014.09.003. URL http://www.sciencedirect.com/science/article/pii/S0893608014002135.

E. Sontag. Nonlinear regulation: The piecewise linear approach. *IEEE Transactions Automatic Control*, 26(2):346–358, 1981.

T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, 15(1):116–132, 1985.

J. Theis, P. Seiler, and H. Werner. LPV model order reduction by parameter-varying oblique projection. *IEEE Transactions on Control Systems Technology*, 26(3):773–784, 2018. URL www.scopus.com.

J. Tomperi, E. Juuso, and K. Leiviskä. Hybrid LE systems for simulation of an activated sludge process. In *Proceedings -8th EUROSIM Congress on Modelling and Simulation, EU-ROSIM 2013, 10-13 September, 2013, Cardiff, UK*, pages 118–123, 2013. doi:10.1109/EUROSIM.2013.31. URL www.scopus.com.

J. Tomperi, E. Juuso, and K. Leiviskä. Early warning of changing drinking water quality by trend analysis. *Journal of Water and Health*, 14(3):433–442, 2016. doi:10.2166/wh.2016.330.

J. Tomperi, E. Juuso, A. Kuokkanen, and K. Leiviskä. Monitoring a municipal wastewater treatment process using a trend analysis. *Environmental Technology (United Kingdom)*, pages 1–10, 2017. doi:10.1080/09593330.2017.1375026.

Y.-W. Xiao, J. Wu, Z.-G. Lin, and X.-D. Zhao. A deep learning-based multi-model ensemble method for cancer prediction. *Computer Methods and Programs in Biomedicine*, 153:1 – 9, 2018. ISSN 0169-2607. doi:10.1016/j.cmpb.2017.09.005.

D.-K. Xu and Y.-J. Tian. A comprehensive survey of clustering algorithms. *Annals of Data Science*, 2(2):165–193, Jun 2015. ISSN 2198-5812. doi:10.1007/s40745-015-0040-1.

L. A. Zadeh. Fuzzy sets. *Information and Control*, 8(June): 338–353, 1965.

EpiRust: Towards A Framework For Large-scale Agent-based Epidemiological Simulations Using Rust Language

Jayanta K. Kshirsagar   Akshay Dewan   Harshal G. Hayatnagarkar

Engineering for Research, ThoughtWorks Technologies, Pune, India,
{jayantak,akshayde,harshalh}@thoughtworks.com

Abstract

To implement large-scale agent-based simulations, developers historically relied on C and C++ due to performance, while struggling to deal with tedious explicit memory management. This struggle translates into software defects and lower developer productivity. More recently, desire to harness multi-core systems via concurrent software complicates design and implementation when memory is shared among compute cores. When we faced this situation, we were looking for a system programming language as fast as C and C++ but without caveats around memory management. Between Go and Rust, we chose Rust language which guarantees safety in memory management even for concurrency, without a run-time or garbage collector. In this paper, we have shared our experience with Rust to build a framework named as EpiRust for agent-based epidemiological simulations. Our simulation experiments have shown some promising results for a population of a few million agents, using commodity-class hardware. Key outcomes of this whole exercise are that EpiRust could achieve following quality goals: 1. Robustness 2. Flexibility 3. Performance.

Keywords: agent-based modeling, agent-based simulations, epidemiology, large-scale simulations, Rust language.

1 Introduction

Pandemics are seen as one of the top global threats, and public health policy experts need tools to assess preparedness and response strategies. Large-scale agent-based epidemiological simulations are one such useful tool in their repertoire (Bisset et al., 2009).

To implement such an epidemiological model, developers can either choose to develop using agent-based modeling platforms such as Netlogo (Tisue and Wilensky, 2004) and GAMA (Taillandier et al., 2019) or using general-purpose libraries and frameworks such as Repast, Flame, and MASON (Rousset et al., 2016).

Precursor to EpiRust was our experience of developing a SEIR model using the GAMA platform. We found that the platform is useful for rapid prototyping and quick feedback on the functionality but performs slower for our needs. The GAMA-based simulations were consuming more time for more number agents in a non-linear fash-

ion. For example, a simulation with 10000 agents would take more than 60 minutes using 128 CPU cores on a large server. We were discouraged with these results as our goal is to simulate a large city like *Pune, India* with more than three million population. These observations triggered us to look for effective alternatives for large scale simulations.

After surveying the literature, we started development of a new framework for large scale epidemiological simulations built on three intrinsic qualities namely, *robustness*, *flexibility*, and *performance*, and we named it as EpiRust.

1.1 Related Work

Developing and maintaining such large-scale simulations is a complex exercise, especially due to a trade-off between flexibility and performance (Rousset et al., 2016; Fujimoto et al., 2017; Chen et al., 2017).

We observed that for large-scale agent-based simulations including epidemiology models, only a few frameworks are designed for this purpose (Abar et al., 2017; Chen et al., 2017). The two important challenges to scale up and out are: 1. Partitioning and distributing the problem across computers, 2. Coordinating the computers to ensure the solution is valid (Yeom et al., 2014).

To develop such distributed systems, choosing a programming language can play an important role in performance, quality and evolution of software. Historically in scientific computing, system programming languages such as C and C++ have been the preferred languages for the superior performance they offer. However, developers using these languages also have low productivity in terms of development time and software quality, and according to (Phipps, 1999), both of these issues emerge from complexity of explicit memory management. In our search for a programming language to develop large-scale epidemiological models, we were looking for the languages that should be fast, stable, with a mature ecosystem of tools and libraries.

C and C++ are designed for system programming which means they can directly interface with underlying hardware. However, Java, C#, and Python languages target respective virtualized runtimes to support hardware-agnostic execution, resulting in lower performance. Julia language (which is designed for scientific computing), and Go language (which is designed for distributed scal-

able applications) adequately fit the bet for large-scale epidemiology simulations. However, both of these languages must rely upon some system programming language to access hardware resources. From a performance point of view, (Kouatchou, 2018) compares many of these languages together to get a sense of their relative strengths and weaknesses.

Software written using C and C++ typically offers better performance however suffers from memory safety and robustness issues due to explicit memory management and concurrency. After evaluating these options, we finally had two choices, Go and Rust language. Between these two, we chose Rust language that guarantees safe memory management without a garbage collector, and the resultant code is safe for concurrency as well (Klabnik and Nichols, 2018; Balasubramanian et al., 2017). Based on similar reasons, a recent work (Antelmi et al., 2019) also has chosen the Rust to implement agent-based simulation models.

1.2 The Rust Programming Language

Rust is a multi-paradigm programming language, which offers features such as zero-cost abstractions, memory safety without garbage collection and run-time overheads, and fearless concurrency.

Rust has a unique approach based on "Ownership and Borrow" principles to free the developer from the burden of manual memory management, an Achilles' Heel for many developers from the perspective of memory safety and robustness in C and C++ (Balasubramanian et al., 2017). Rust maintains only one reference to the allocated memory eliminating the need of garbage collection and resulting in faster execution performance. All the variables being non-mutable by default, Rust compiler forces a developer to write smaller and cleaner code. For flexibility and testability, Rust prefers composition over inheritance via traits.

In this paper, we have shared our rationale and learning of implementing a large-scale epidemiological simulation scaling up to a population of a few million agents and capable of running on commodity-class computer hardware in reasonable time. In section 2 discusses the current EpiRust model including population, geography, disease dynamics and interventions. In the following section 3, we have discussed the features of the Rust language, their relevance in the EpiRust implementation, and later the results of simulating baseline and lockdown scenarios for the city of Pune. Finally, we have concluded the paper with our observations and plans for the EpiRust framework.

2 EpiRust Model

The EpiRust model follows a minimalist approach for modeling an epidemic in a virtual society along three main aspects namely *population*, *geography*, and *disease*.

2.1 Population

An agent represents a person in a population. Agents are heterogeneous along two attributes namely *work status* (via *working* and *non-working*) and *transport preference* for office commute (via *public transport* and *private transport*). A fraction of the working population can be configured as the *essential workers* whose services are available even when the other agents have restrictions on their movements. Each agent follows a daily routine, prescribed by its demographics and other attributes, as shown in table 1. An agent could catch and spread infection while interacting with other agents based on how contact probabilities change due to their routine.

2.2 Geography

Factors such as geography, demographics, and transportation influence the spread and the speed of epidemics. The EpiRust model mimics geography of a minimalist city and represents it as a *grid* with various planned functional areas for homes, workplaces and transportation.

- **Home Area** Every agent is randomly assigned a home location during initialization. Collectively, these home locations represent a residential colony. Working agents commute to-and-fro their workplaces, whereas non-working agent movements are restricted to this area.

- **Public Transport Area** is used by working agents for home and work commute. Currently, the model assumes that agents using private transport reach directly to homes and workplaces.

- **Work Area** During work hours, a working agent occupies a dedicated work location in this area with restricted movements.

Figure 1. Grid-based Geography Depiction

Start Hour	End Hour	Start Place	End Place	Activity
0	8	Home	Home	No movement (Sleep, etc.)
8	9	Home	Home	Activity within home
9	10	Home	Transport	Go to public transport
10	11	Transport	Transport	Move within transport
11	12	Transport	Work	Go to work
12	19	Work	Work	Move within work
19	20	Work	Transport	Go to public transport
20	21	Transport	Home	Go to home
21	23	Home	Home	Activity within home

Table 1. 24-hour Routine of Working Agents

2.3 Disease Dynamics

The current disease dynamics is based on the Mordecai SEIR model (Childs et al., 2020), and is specialized for COVID-19 similar to (Snehal Shekatkar et al., 2020). The model is shown in figure 2, and the disease progression over time is rendered in figure 3.

1. **Susceptible** All agents are initialized in the susceptible compartment.

2. **Exposed** A susceptible agent could get exposed to an infection if it comes in contact with an infectious agent. An exposed agent does not infect other susceptible agents immediately. It stays in this state for an incubation period specific to a disease. After this period, the agent state would switch to the infected compartment. While initializing the simulation, a fraction of the population can be configured as being in the exposed compartment.

3. **Infected** In this state, an agent could infect other susceptible agents with some probability. This probability is specific to a disease, and can be configured by a user. An infected agent could either be symptomatic or asymptomatic. Symptomatic agents could later be mild or severely infected. On the other hand, the asymptomatic agents would not exhibit any symptoms but still could infect others.

4. **Removed** The recovered and the deceased agents would belong to this state. A SEIR model assumes that the recovered agents would not catch an infection again in their lifetime. The death rate is an input parameter to decide an agent's recovery.

2.4 Interventions

Interventions play a key role in devising policies to reduce and eventually to stop the spread of an epidemic. An intervention can be specific to demographic groups within a population or their actions. EpiRust has support for three interventions namely isolation via *hospitalization*, *lockdown* of the city, and *mass vaccination*.

- **Hospitalization** intervention helps to isolate and quarantine agents with severe infection. Hospitalization is limited by the number of hospital beds, but this limit can be specified via configuration parameters.

- **Lockdown** would restrict citizen movements. Agents would be confined in their homes and only essential workers would be allowed to travel.

- **Mass Vaccination** strategy vaccinates user-defined percentage of randomly chosen susceptible agents. Once vaccinated, these agents would never catch disease in their lifetime. This intervention can be triggered by a configurable threshold.

2.5 Simulation Loop

At the heart of the simulation, there are two simple loops: The outer one controls the discrete time steps (or clock), and the inner loop updates the state of each agent at each time step.

```
for hour in 1..n do
    if can_intervene then
        │ Apply lockdown;
    end
    for each agent do
        │ Move agent on the grid;
        │ Update infection state;
    end
    if number_of_infected == 0 then
        │ Stop simulation;
    end
end
```
Algorithm 1: Simplified Simulation Loop

Here, updating an agent after another in a sequential execution leads to path dependent outcomes. To solve this issue, EpiRust uses a double buffering technique (Cosenza et al., 2018; de Aledo Marugán et al., 2018). The technique employs a read-only buffer for referring to the current state of agents, and a write-only buffer for updating their next state.
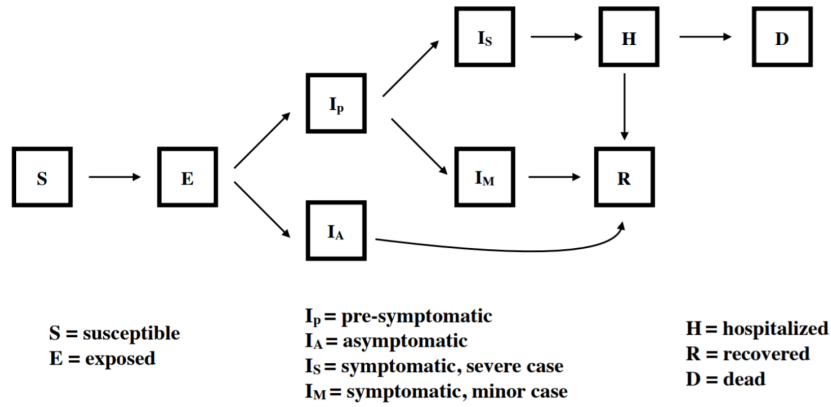
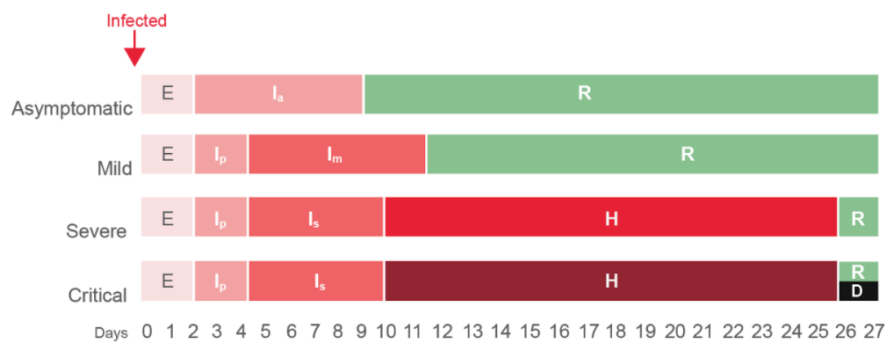**Figure 2.** The Mordecai SEIR Model (Childs et al., 2020)

Figure 3. Disease Dynamics (Snehal Shekatkar et al., 2020)

3 EpiRust Implementation

As we discussed earlier, EpiRust has been designed with three goals: Robustness, flexibility, and performance. Rust contributes to robustness of memory management and concurrent access via its features of borrow checking and lifetime management (Klabnik and Nichols, 2018). In the rest of this section, we have discussed the implementation towards flexibility and performance goals.

3.1 System Requirements

3.1.1 Model Correctness

To verify the model, a basic grid visualization is developed to observe the agents during each time step of the simulation. Along with the grid visualization, we followed *test-driven development* (*TDD* for short) to detect defects and similar regression caused by code changes. TDD is part of *Agile Software Development* practices, where a developer writes unit test cases before writing the functional code, and runs these test cases after implementing a change (Beck, 2003). This approach helps in verifying functional correctness, especially in case of the frequent code changes. Cargo (Klabnik and Nichols, 2018), Rust's build and package manager gives an ability to run unit tests and performance benchmark tests. Unit tests rely upon *Mocks* which mimics structure and behavior of code beyond the scope of the current unit. We found that creating *mocks* for Rust structures is not convenient for developers.

3.1.2 Flexibility

A model is flexible if a user can simulate different variations of the simulation model, either by specifying configuration parameters or disease-specific model extensions.

EpiRust uses the Rust constructs such as *mod*ules for modularity, *struct*ures for entity modeling, *trait*s to define behavioral contracts for later composition with entities (something similar to inheritance in other languages). These constructs have contributed to separate disease dynamics from other components. This separation enables the necessary *flexibility* to attach any SEIR-based disease dynamics to the simulation without significant changes in the rest of the source code. A user can alter disease dynamics by altering configuration parameters for not only COVID19 but also other SEIR diseases such as smallpox. For example, to model smallpox disease dynamics, a user can choose to not specify asymptomatic and mild symptomatic transmission rates in the configuration parameters.

The user can simulate multiple scenarios without changing source code via configurable input parameters. A subset of these parameters are discussed below.

General Parameters

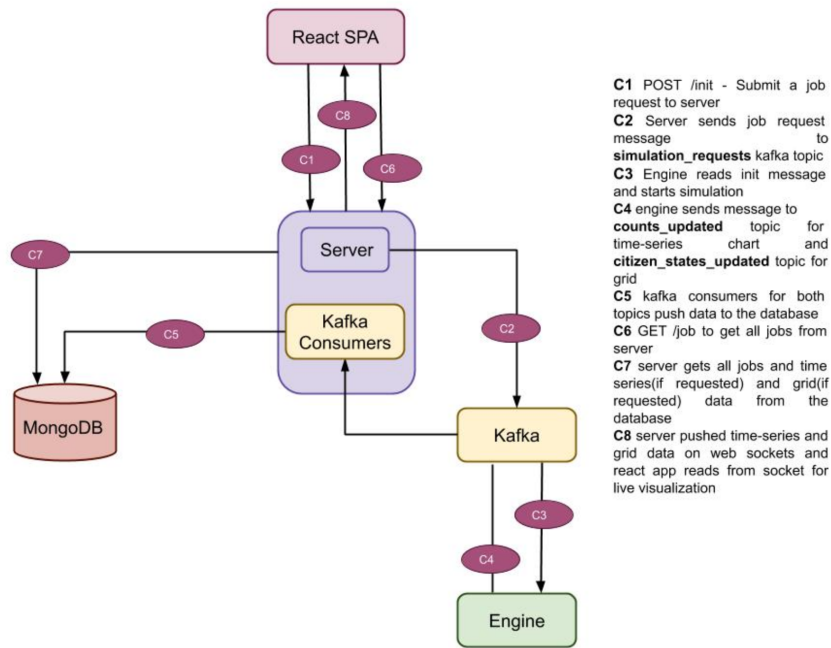- **grid_size** size of the square grid to represent the city.

Figure 4. System Architecture

- **number_of_agents** the population in the city.

- **working_percentage** the percentage of working population to total population.

- **public_transport_percentage** the percentage of the population taking public transport to work.

- **hospital_beds_percentage** the number of hospital beds per 100 people.

- **starting_infections** the seed infected population.

Disease Dynamics Parameters

- **percentage_asymptomatic_population** specifies the percentage of population will not show any symptoms when infected.

- **percentage_severe_infected_population** Fraction of the infected population with severe symptoms would need hospitalization.

- **transmission_start_day** Day after catching an infection when an infected person may spread the disease.

- **transmission_rate** is the probability of a susceptible agent catching an infection when contacted by an infected agent.

- **asymptomatic_last_day** is the last day for an asymptomatic agent to recover completely from an infection.

- **last_day** is the last day of the disease when an agent either recovers completely or dies of a severe infection.

Intervention Parameters

- **Lockdown.at_number_of_infections** triggers the lockdown intervention when the total number of infected agents matches to the parameter value.

- **Lockdown.essential_workers** is the fraction of the working population which would perform their work routine even during lockdown.

- **Vaccination.at_hour** is the time step to trigger vaccination.

- **Vaccination.percent** is a fraction of the population to be randomly vaccinated.

The EpiRust disease dynamics model works as a finite state machine (FSM) with the states as the SEIR compartments. For state transitions, FSM first matches the current state and then moves to the next state. To ease this implementation, Rust offers a construct *Pattern* for matching numbers, strings, *enum*s, and regular expressions, etc.

3.1.3 Scalability

While representing millions of agents with their movements and disease states along with the geographical details CPU and memory consumption are two important factors to optimize. Choosing appropriate data structures and algorithms can help to optimally manage these resources. For example, when the model used a dense matrix for representing the grid-based geography for agent locations, it consumed a large amount of memory throughout the simulation for even smaller agent populations. An efficient alternative is spatial hashing technique which

stores only the locations where agents are placed. Switching from the dense matrix to the spatial hashing reduced memory usage to a fraction of its original.

3.1.4 Performance

While spatial hashing helped to reduce memory usage, it kept CPU time consumption at the same level of dense matrix. For the hashmap data structure, the average-case time complexity for retrieval and insertion is $O(1)$. After evaluating multiple implementations of hashmap data structure in the Rust ecosystem, EpiRust uses *FNVHash* (Crichton, 2020) for better performance. In addition, we observed that the random number generator supplied by the Rust standard library is adequate to avoid collisions during agent placement.

3.2 Architecture

EpiRust has multiple smaller components as depicted in figure 4. **Engine** is the simulation engine written in *Rust* and discussed in this paper. The engine is invoked via command line inputs including configuration files in the JSON format. **Server** is a web server and the central component which interfaces with other components such as the engine and the user interface. It invokes the engine using its command line interface. It sends simulation requests to the engine and reads the engine output. The output is later stored in a database so as to consume it later. The server uses *Apache Kafka* for communication with the engine. Kafka is a distributed streaming platform used here for mediation between the server and the engine. **MongoDB** stores simulation-generated data for later consumption for visualization and debugging. **React SPA** is the web user interface to specify model parameters and visualize the simulation results. It has been developed as a single page application using *React JS*.

4 Scenarios and Results

When EpiRust development had started, COVID-19 had begun to unfold in *Wuhan, China*. In order to get a better understanding of COVID-19 spread in real life, we decided to run it for the city of *Pune, India*. The city is spread over an area approximately *331 square km* in the Western India, and has a population of more than three million people.

To represent Pune city, we provided the following configuration parameters as rendered in table 2.

To get a better understanding of the impact of various non-pharmaceutical interventions on spread of a disease, we simulated two different scenarios namely *Baseline* and *Intervention*. As a model is probabilistic, we ran each of these scenarios 5 times to ensure statistical stability.

4.1 Baseline Scenario

The baseline scenario depicts business as usual even in a pandemic scenario. All agents will follow their routing regardless of the disease spread and all of the interventions

Parameter	Value
number_of_agents	3137224
grid_size	5660
working_percentage	70%
public_transport_percentage	80%
Lockdown_at_number_of_infections	100
essential_workers_population	10%
hospital_beds_percentage	0.003%
starting_infections	32645

Table 2. Input Parameters

like lockdown, increasing hospital bed capacity, vaccination etc. would not be in place.

4.2 Intervention Scenario

Lockdown can be a good intervention strategy but has its own costs. To balance cost versus benefits requires trying various what-if scenarios. EpiRust lockdown intervention can be triggered for a configurable threshold of number of infections. During lockdown intervention, the essential workers would still be working and performing their daily routine as discussed in 2.1, and could potentially acquire and spread infections. A fraction of the population could choose to defy the lockdown intervention. This non-compliant population could be modeled by adding their percentage to the essential workers. Mass vaccination as an intervention has been turned off for these simulations.

4.3 Results

We used Rust compiler version 1.44.1 for all these benchmarks, and each scenario of this stochastic model has been run for at least five times to ensure statistical stability. So, the epicurves charts 5, 6(a) and 6(b) render mean and variance over time.

Compared to a similar model implemented using the GAMA platform, a single-threaded EpiRust simulation now completes within 30 seconds (a speedup of more than 100x). For Pune city simulations with more than 3 million population, EpiRust takes close to 150 minutes on single core of commodity class hardware for baseline scenario and 50 minutes for lockdown scenario.

For every agent in the simulation, the number of interactions with neighboring agents and its associated time complexity are order of $O(n^3)$. We observed that the Rust's *HashMap* based implementation has a smaller memory footprint, just over twenty megabytes for a million agents, and hence computer memory is less of a concern here. The performance across processors varies due to clock frequency as well as internal architecture of the processor.

While comparing multiple simulation scenarios for Pune city, we found that lockdown is an effective strategy to reduce the total infections in the society. Based on our observations (figures 6(a), 6(b)) lockdown slows down the rate at which infections grow and imposing lockdown
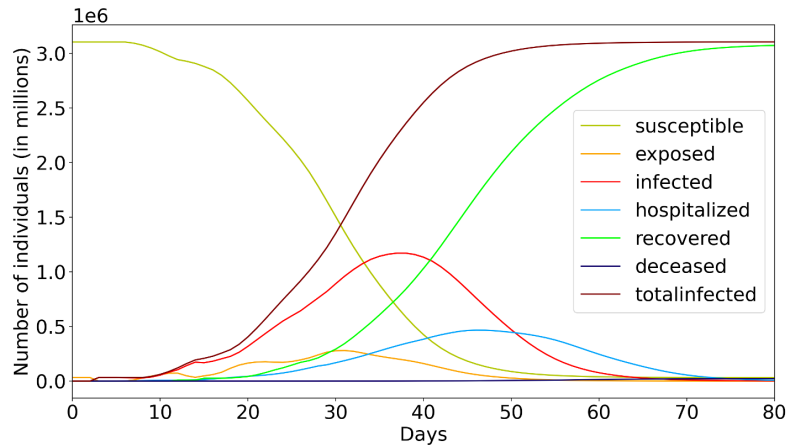
Figure 5. Pune Baseline Scenario

(a) Lockdown

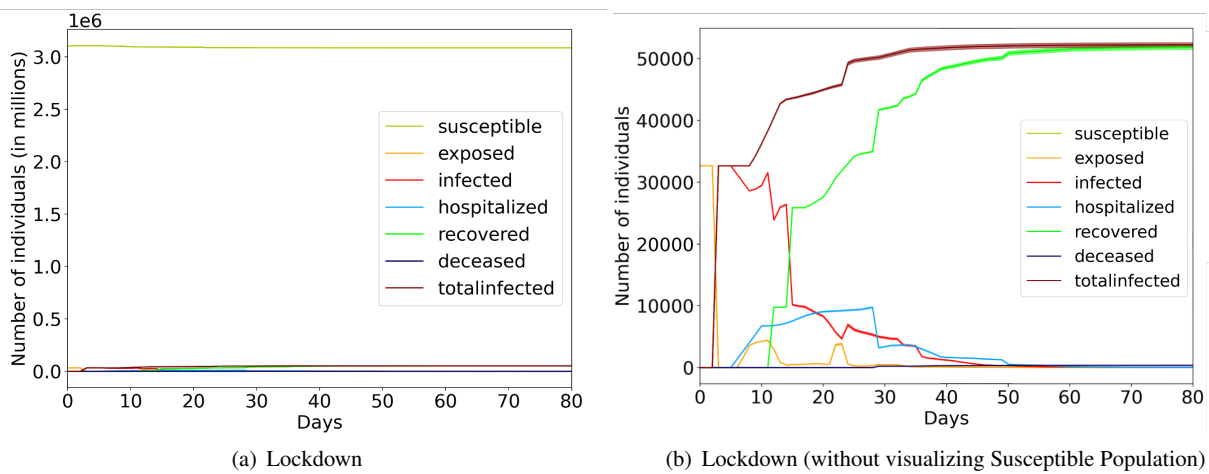(b) Lockdown (without visualizing Susceptible Population)

Figure 6. Lockdown Scenarios for Pune Simulations

reduces stress upon health care facilities like hospitals.

4.4 Reproducing the results

EpiRust source code is released in open source. To reproduce results of this paper, the EpiRust framework and the model used in this paper can be found here: `https://github.com/thoughtworks/epirust/tree/sims2020`. In cloned Git repository, the configuration files used for baseline and lockdown scenarios can be found at `epirust/engine/config/high_baseline.json` and `epirust/engine/config/high_lockdown.json` respectively.

5 Conclusion and Future Work

Based on our current exploration, we found Rust useful in writing large-scale epidemiological models, even without employing some advanced features such as parallelism and concurrency. Rust's preference for composition over inheritance helps in building flexible models. We found that Rust has a steeper learning curve than other programming languages, especially to interpret the compiler errors around borrow-checking. However, we trust that once

compiled the generated executable is robust and guaranteed to be memory safe. As a bonus, we found the ownership principle guarantees efficient and memory safe code even for concurrency. The Rust package manager *Cargo* handles library dependencies and their installation transparently. Rust's borrow checker expects the programmer to specify the lifetime of the function parameters to check the reference validity. This can make writing extensible code to be tricky. We observed that writing unit tests for complex nested structures is difficult due to limited support for mocking.

Using the Rust and its ecosystem, we believe our success is moderate in achieving our goals of performance, flexibility, and robustness qualities. At the same time, this lays the foundation of our future work on this framework. For example, to improve performance, we would like to harness multiple cores across computers. We also intend to explore features for flexibility around dynamic and user-defined disease models. We acknowledge the impact of demographics on simulation results, and will work towards supporting synthetic population. To ease setup and deployment, we are working on a Docker image and

will be shared soon.

6 Acknowledgements

References

Sameera Abar, Georgios K Theodoropoulos, Pierre Lemarinier, and Gregory MP O'Hare. Agent based modelling and simulation tools: A review of the state-of-art software. *Computer Science Review*, 24:13–33, 2017.

Alessia Antelmi, Gennaro Cordasco, Matteo D'Auria, Daniele De Vinco, Alberto Negro, and Carmine Spagnuolo. On Evaluating Rust as a Programming Language for the Future of Massive Agent-Based Simulations. In *Asian Simulation Conference*, pages 15–28. Springer, 2019.

Abhiram Balasubramanian, Marek S. Baranowski, Anton Burtsev, Aurojit Panda, Zvonimir Rakamaric, and Leonid Ryzhyk. System programming in rust: Beyond safety. *Operating Systems Review*, 51:94–99, 2017.

Kent Beck. *Test-driven development: by example*. Addison-Wesley Professional, 2003.

Keith R Bisset, Xizhou Feng, Madhav Marathe, and Shrirang Yardi. Modeling interaction between individuals, social networks and public policy to support public health epidemiology. In *Proceedings of the 2009 Winter Simulation Conference (WSC)*, pages 2020–2031. IEEE, 2009.

Jiangzhuo Chen, Bryan Lewis, Achla Marathe, Madhav Marathe, Samarth Swarup, and Anil KS Vullikanti. Individual and collective behavior in public health epidemiology. In *Handbook of statistics*, volume 36, pages 329–365. Elsevier, 2017.

M. L. Childs, M. Kain, D. Kirk, M. Harris, L. Couper, N. Nova, I. Delwel, J. Ritchie, and E. Mordecai. The impact of long-term non-pharmaceutical interventions on covid-19 epidemic dynamics and control. *medRxiv*, 2020.

Biagio Cosenza, Nikita Popov, Ben H. H. Juurlink, Paul Richmond, Mozhgan Kabiri Chimeh, Carmine Spagnuolo, Gennaro Cordasco, and Vittorio Scarano. Openabl: A domain-specific language for parallel and distributed agent-based simulations. In *Euro-Par*, 2018.

Alex Crichton. fnv hash. https://crates.io/crates/fnv/1.0.7, 2020. URL https://crates.io/crates/fnv/1.0.7.

Pablo González de Aledo Marugán, Andrey Vladimirov, Marco Manca, Jerry Baugh, Ryo Asai, Marcus Kaiser, and Roman Bauer. An optimization approach for agent-based computational models of biological development. *Adv. Eng. Softw.*, 121:262–275, 2018.

Richard M Fujimoto, Christopher Carothers, Alois Ferscha, David Jefferson, Margaret Loper, Madhav Marathe, and Simon JE Taylor. Computational challenges in modeling & simulation of complex systems. In *2017 Winter Simulation Conference (WSC)*, pages 431–445. IEEE, 2017.

Steve Klabnik and Carol Nichols. *The Rust Programming Language*. Mozilla Research, 2018.

J Kouatchou. Basic comparison of python, julia, matlab, idl and java (2018 edition). *Modeling Guru–National Aeronautics and Space Administration, USA*, 2018.

Geoffrey Phipps. Comparing observed bug and productivity rates for Java and C++. *Software: Practice and Experience*, 29(4):345–358, April 1999. ISSN 0038-0644, 1097-024X.

Alban Rousset, Bénédicte Herrmann, Christophe Lang, and Laurent Philippe. A survey on parallel and distributed multi-agent systems for high performance computing simulations. *Computer Science Review*, 22:27–46, November 2016. ISSN 15740137.

Snehal Shekatkar, Bhalchandra Pujari, Mihir Arjunwadkar, Dhiraj Kumar Hazra, Pinaki Chaudhuri, Sitabhra Sinha, Gautam I Menon, Anupama Sharma, and Vishwesha Guttal. Indscisim a state-level epidemiological model for india, 2020. Ongoing Study at https://indscicov.in/indscisim.

Patrick Taillandier, Arnaud Grignard, Nicolas Marilleau, Damien Philippon, Quang-Nghi Huynh, Benoit Gaudou, and Alexis Drogoul. Participatory modeling and simulation with the gama platform. *Journal of Artificial Societies and Social Simulation*, 22(2):3, 2019. ISSN 1460-7425. URL l.

Seth Tisue and Uri Wilensky. Netlogo: A simple environment for modeling complexity. In *International conference on complex systems*, volume 21, pages 16–21. Boston, MA, 2004.

Jae-Seung Yeom, Abhinav Bhatele, Keith R. Bisset, Eric J. Bohm, Abhishek Gupta, Laxmikant V. Kalé, Madhav V. Marathe, Dimitrios S. Nikolopoulos, Martin Schulz, and Lukasz Wesolowski. Overcoming the scalability challenges of epidemic simulations on blue waters. *2014 IEEE 28th International Parallel and Distributed Processing Symposium*, pages 755–764, 2014.

Automated synthesis of netlists for ternary-valued n-ary logic functions in CNTFET circuits

Halvor Nybø Risto   Steven Bos   Henning Gundersen

Ternary Research Group, Department of Science and Industry Systems, University of South-Eastern Norway, Norway, `{henning.gundersen}@usn.no`

Abstract

This paper is an investigation of automated netlist synthesis for ternary-valued n-ary logic functions, based on a static ternary gate design methodology. We present an open-source C++ implementation, which outputs a ready-to-simulate SPICE subcircuit netlist file for ternary-valued n-ary function circuits. A circuit schematic of the 3-operand carry is demonstrated as synthesized by the netlist generator. We investigate a holistic (non-compound) approach to designing balanced full-adders by using 3-operand func-tions as compared to a traditional 2-operand compound design methodology. Three gate-level design approaches (compound, non-compound and hybrid) for the balanced full-adder have been simulated in HSPICE and are com-pared to each other and the state-of-the-art with simulation results. Furthermore, we propose to standardize the ternary functions by indexing them. This indexing system allows for the convenience of referencing any possible logic func-tion with no ambiguity. This indexing is necessary as most ternary functions do not have semantic names (e.g. AND, OR) and the amount of unique 3-valued functions grows exponentially with higher arity.

Keywords: ternary, netlist, synthesis, simulation

1 Introduction

In recent years, along with developments of the carbon nanotube field-effect transistor (CNTFET), there have been a handful of papers on the designs and synthesis of ternary or 3-valued logic gates implemented in simulations of CNTFET circuits. One paper in particular (Kim et al., 2018) proposes a design method for ternary logic gates, with the use of pull-up and pull-down networks constructed from a truth table for the circuit. In ternary logic, with only two operands, there are 19683 possible logic gates. With three operands, 7.6e12 logic gates are possible. The process of designing the circuit and writing the netlists of these circuits can be a tedious process, especially for circuits with more than two operands. Therefore, an open-source netlist synthesizer is of much use. The study (Lee et al., 2019) reports to have automated this process, however their code is not open-source.

2 Function Indexing

To unambiguously refer to any of the many logic functions, we propose a simple indexing system.

2.1 Range of index in arities

The number of possible functions for a specific arity and radix can be calculated as in Equation 1, where R is the radix and A is the arity.

$$F_{range} = R^{R^A} \qquad (1)$$

Table 1. Range of functions in arities and radices

Arity	Radix 2	Radix 3
1	$2^{2^1} = 4$	$3^{3^1} = 27$
2	$2^{2^2} = 16$	$3^{3^2} = 19683$
3	$2^{2^3} = 256$	$3^{3^3} = 7,625,597,484,987$

While in binary, there are few enough functions that naming the useful functions (AND, OR, XOR, etc.) has been a feasible practice, for ternary logic the quantity of possible functions is more unwieldy, as can be seen in Table 1. Therefore, an indexing system is proposed to refer to specific ternary-radix functions.

The indexing system maps every truth table to an index by counting up from 0 to the function range, along with the values of the truth table. As an example, a function always outputting the low value would be the 0th index. Then, the first row of the truth table acts as the three lowest-significance trits of the index, and so on. With a truth table listed vertically, the output values for a specific function can be read in ternary as the function index.

2.2 Heptavintimal index encoding

We adopt the usage of the base-27 heptavintimal notation for ternary values (Jones, 2012), as it conveniently covers three ternary digits (trits) per symbol, as shown in Table 2. As one operand can have one of three values, the truth table of a function is three trits long in each dimension. Therefore, a logic function index can conveniently be encoded with the heptavintimal notation.

Table 2. The heptavintimal notation

Weight(Decimal)	0	1	2	3	4	5	6	7	8	9	10	11	12	13
Ternary	000	001	002	010	011	012	020	021	022	100	101	102	110	111
Heptavintimal	0	1	2	3	4	5	6	7	8	9	A	B	C	D

Weight(Decimal)	14	15	16	17	18	19	20	21	22	23	24	25	26	
Ternary	112	120	121	122	200	201	202	210	211	212	220	221	222	
Heptavintimal	E	F	G	H	K	M	N	P	R	T	V	X	Z	

3 Methodology

Based on the static ternary gate design methodology of (Kim et al., 2018), we have implemented an algorithm in C++, which produces a ready-to-simulate SPICE subcircuit netlist file from a circuit truth table.

The program takes an n-dimensional truth table, and constructs the four truth tables for the pull-up and pull-down networks. Then, for each network, a set of n-dimensional rectangular groupings are found. Each of these groupings will provide a transistor path to the output within each pull-up and pull-down network.

3.1 Usage

To use the program, compile the open-source code in a C++ compiler. The netlists will be generated in the same file directory as the compiled program. When the program starts, it asks for the function arity, and the values of each element in the three-by-three n-dimensional truth table, with values low(0), middle(1), high(2), don't care(x). The filename will be generated as the function index of the specific function. The transistor parameters, as well as which CNTFET model is being used, can be specified with the string variables p0, p1, p2, n0, n1, n2.

The program will produce a subcircuit which must be connected externally to a 0.9V voltage supply, and the operand inputs. The circuits rely on external 2-transistor Positive Ternary Inverters (PTI) and Negative Ternary Inverters (NTI) to achieve the four different transistor operations detailed in (Kim et al., 2018).
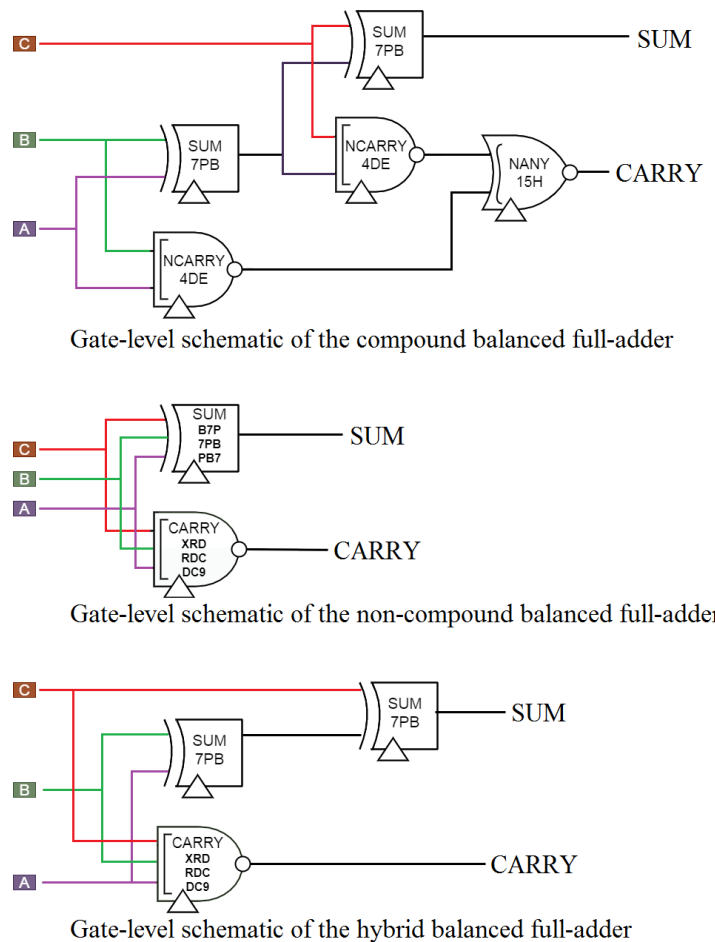
3.2 Logic minimization algorithm

The logic minimization done to produce an optimized circuit is similar to karnaugh-mapping. The grouping algorithm takes the truth tables for each transistor network and draws n-dimensional rectangular groupings which covers every '1' on the truth table for each network, with as few groupings as possible. These groupings represent the transistor-paths in the circuit towards the output within each of the four transistor networks. Each transistor in series narrows down the throughput, until the logical rectangle of a grouping is achieved.

4 Circuit schematics

The common procedure for constructing functions with more than two operands is to combine smaller functions to create bigger compound functions. However, circuits with more operands can also be generated with our program. Therefore we investigate the usage of 3-operand functions in a 1-trit balanced full-adder circuit, in the form of a non-compound and a hybrid gate architecture. These architectures are a more holistic view of the function as a relation between input and output. Figure 1 shows three different balanced full-adder circuit design approaches.

Gate-level schematic of the compound balanced full-adder

Gate-level schematic of the non-compound balanced full-adder

Gate-level schematic of the hybrid balanced full-adder

Figure 1. Three approaches for a balanced full-adder

Figure 2 shows the circuit schematic for the 3-carry circuit used in the hybrid 1-trit full-adder, with diameters 1.487 *nm* and 1.018 *nm* being depicted as blue and green respectively.

Table 3. Simulation results with 2fF load capacitor

Circuit	Transistors	Avg. power 500MHz	Avg. power 50MHz	Worst Delay	PDP 500MHz	PDP 50MHz
2-sum (7PB)	40 (32)	$0.61\mu W$	$0.35\mu W$	530ps	0.327e-15 J	0.185e-15 J
2-ncarry (4DE)	10	$0.80\mu W$	$0.80\mu W$	20ps	0.016e-15 J	0.016e-15 J
2-nany (15H)	18	$0.33\mu W$	$0.32\mu W$	40ps	0.013e-15 J	0.013e-15 J
3-sum (B7P7PBPB7)	150 (138)	$0.56\mu W$	$0.34\mu W$	1530ps	0.856e-15 J	0.526e-15 J
3-carry (XRDRDCDC9)	50 (38)	$0.87\mu W$	$0.82\mu W$	30ps	0.026e-15 J	0.024e-15
(Lee et al., 2019) Unbalanced FA	106	(no data reported)	$0.47\mu W$	0.89ns	(no data reported)	**0.421e-15 J**
(Vudadha et al., 2018) Unbalanced FA	98	(no data reported)	**$0.43\mu W$***	1.57ns*	(no data reported)	0.667e-15 J*
Proposed Compound Balanced FA	118 (102)	$2.73\mu W$	$2.29\mu W$	**0.55ns**	1.44e-15 J	1.262e-15 J
Proposed Non-compound Balanced FA	188 (176)	**$1.67\mu W$**	$1.18\mu W$	1.53ns	2.55e-15 J	1.805e-15 J
Proposed Hybrid Balanced FA	118 (102)	$1.96\mu W$	$1.50\mu W$	0.56ns	**1.10e-15 J**	0.840e-15 J

* see (Lee et al., 2019)
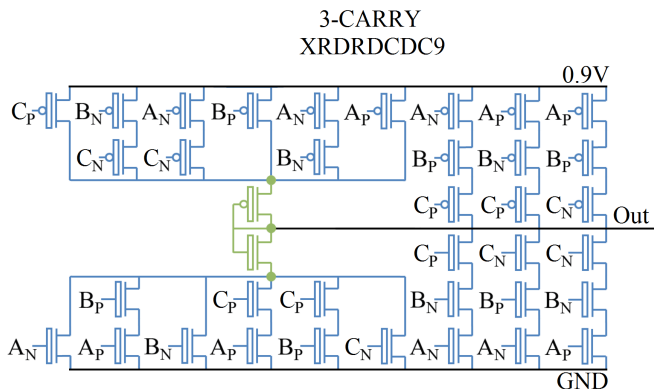
3-CARRY
XRDRDCDC9

Figure 2. The balanced 3-operand carry function

5 Simulation results

The simulations were done in HSPICE, with the standard 32nm CNFET model technology from Stanford University. (Deng and Wong, 2007)

For the sake of these simulations, voltages below 200mV is considered "low", 250mV to 650mV is "middle", and above 700mV is "high".

Table 3 shows simulation results for some balanced functions, and compares three different circuit concepts of a balanced full-adder. The average current is measured at 500MHz and 50MHz. All measurements include the external PTI and NTI inverters of 2 transistors each where they are required. The transistor count is shown with and without the external inverters. A capacitive load of 2fF was put on the output to ground.

6 Discussion

The runtime performance of the synthesizer can be further optimized for > 7 arity. Under that condition circuit solutions can be found in reasonable time on standard hardware. Due to the sheer number of possible functions, only a minority of the circuit solutions were tested. However, all the tests produced the correct output values.

It should be possible to optimize the circuit solutions even further as we found by manual inspection. This is especially true for circuits with high arity functions. It is interesting to see that the performance of a 1-trit balanced ternary full-adder compared is comparable to an unbalanced version, commonly found in literature.

7 Conclusion

For up to 7 operands, a circuit of any 1-output ternary-valued function can be produced. We show that 3-operand functions can be implemented in circuits such as a 1-trit balanced full-adder, and may in some cases outperform a traditional 2-operand design strategy, as the hybrid full-adder was shown to outperform the compound full-adder in terms of power-delay-product (PDP) performance.

This paper has provided three contributions:

1. An open-source implementation for synthesis of n-ary ternary-valued CNTFET circuits (Risto, 2020).

2. An indexing system has been proposed which allows for any possible ternary-valued logic function to be referenced unambiguously.

3. A novel 3 operand, classical 2 operand, and a hybrid 1-trit balanced full-adder circuits have been simulated and compared with simulation results.

References

J. Deng and H. . P. Wong. A compact spice model for carbon-nanotube field-effect transistors including nonidealities and its application—part i: Model of the intrinsic channel region. *IEEE Transactions on Electron Devices*, 54(12):3186–3194, 2007.

Douglas W. Jones. The Ternary Manifesto · Heptavintimal encoding of ternary values, 2012. URL http://homepage.divms.uiowa.edu/~jones/ternary/hept.shtml.

S. Kim, T. Lim, and S. Kang. An optimal gate design for the synthesis of ternary logic circuits. In *2018 23rd Asia and South Pacific Design Automation Conference (ASP-DAC)*, pages 476–481, 2018.

S. Lee, S. Kim, and S. Kang. Ternary logic synthesis with modified quine-mccluskey algorithm. In *2019 IEEE 49th International Symposium on Multiple-Valued Logic (ISMVL)*, pages 158–163, 2019.

Halvor Nybø Risto. Automated synthesis of netlists for ternary-valued n-ary logic functions in cntfet circuits, September 2020. URL https://doi.org/10.5281/zenodo.4015574.

C. Vudadha, A. Surya, S. Agrawal, and M. B. Srinivas. Synthesis of ternary logic circuits using 2:1 multiplexers. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 65 (12):4313–4325, 2018.