SIMS 2023

Proceedings of the 64th International Conference of Scandinavian Simulation Society

Västerås, Sweden

September 25-28, 2023

Editors:
Prof. Konstantinos G. Kyprianidis
Prof. Erik Dahlquist
Dr. Ioanna Aslanidou
Dr. Avinash Renuke
Dr. Gaurav Mirlekar
Prof. Tiina Komulainen
Prof. Lars Eriksson

Published by:
Linköping University Electronic Press (LiU E-press)

Linköping Electronic Conference Proceedings, No 200

SPONSORS


Mälardalen
University


LINKÖPING
UNIVERSITY


Automation
Region


Electrification
Hub


Tekniska Föreningen
med Elektriska Klubben

Preface

We are very pleased to introduce the proceedings of the 64th International Conference of Scandinavian Simulation Society, SIMS2023, held at The Steam Hotel in Västerås, Sweden, during the period September 25–28, 2023.

SIMS2023 is open to scientific contributions from broad aspects of recent research and development work in modelling, simulation, and optimisation in engineering applications. This includes modelling and simulation for design, planning, optimisation, control, and monitoring as well as practical case studies of industrial automation. Furthermore, tools and toolboxes for modelling and simulation, numerical methods for simulation, novel modelling techniques, and visualization of modelling and simulation results are of interest. Typical applications include but are not limited to (i) Renewable energy systems (bioenergy and biofuels, geothermal, hydro, solar, thermal, wave, tidal, and wind energy), (ii) Hydrogen technologies (production, storage and transportation, hydrogen value chain), (iii) Energy systems (electric power, energy storage, fuel cells, heat pumps, industrial plants, energy use in buildings, power plants), (iv) Transportation (automotive, hybrid and electrical vehicles, marine, infrastructure), (v) Industrial processes (carbon capture and storage, chemical processing, oil and gas, and water treatment, (vi) Cyber-physical systems, and (vii) Biosystems and medical systems.

The scientific program of SIMS2023 was curated by an international program committee, consisting of researchers in simulation, artificial intelligence and machine learning, energy systems and industrial technology. The main goal was to deliver a high-quality program with a wide variety of topics. In the single-blind review process, at least three reviewers evaluated each submission and gave their recommendations. Ambiguous results were thoroughly discussed among the reviewers and the senior members of the program committee, who then made the final selection.

SIMS2023 is the first conference on simulation of its kind in Europe starting as early as 1959, and to the best of our knowledge, it is the only conference on simulation that covers all aspects related to simulation in a diverse range of applications. The program committee welcomed submissions in the following formats: ***regular technical papers*** up to 8 pages (including references) on substantial, original, and unpublished research, including evaluation results, where appropriate, and ***poster*** submissions on focused contributions from students and industries and ***tutorials***. In addition, we were very fortunate to have two high-profile keynote speakers: ***Dr. Alf Isaksson*** (ABB, Sweden), an expert on Automation and Industrial Autonomy; and ***Dr. Jianping Wang*** (Hitachi Energy Research, Sweden), a renowned expert on Power Systems protection.

These proceedings include only the regular technical papers that were papers accepted by the review committee and were presented during the conference. In more detail, the conference received *77* submissions from all over the world, including from China, Germany, Finland, Norway, France, Germany, and Italy. We were able to accept ***52*** full papers, ***8*** posters and ***2*** tutorials. ***51*** accepted and presented papers are collected in this volume.

Organizing a conference and a peer-review process always relies on the good will and support of many colleagues who take their valuable time and contribute to an interesting and fruitful program. Firstly, we would like to thank the members of the conference organising committee for putting a great technical program.

Our special thanks go to all the technical area chairs of the technical program committee, Prof. Erik Dahlquist (Mälardalen University), Dr. Valentina Zaccaria (Mälardalen University), Dr. Amare Desalegn Fentaye (Mälardalen University), Prof. Mirko Morini (University of Parma), Dr. Valentin Scheiff (Mälardalen University), Dr. Hao Chen (Mälardalen University)., Prof. Eva Thorin (Mälardalen University), Dr. Stavros Vouros (Mälardalen University), Dr. Gaurav Mirlekar (University of South-Eastern Norway), Dr. Amir Vadiee (Mälardalen University) and Dr. Avinash Renuke (Mälardalen University), for their substantial support and for many constructive online meetings. It was a pleasure working with you. I also want to thank our industrial program committee, Dr. Lokman Hosain (Hitachi-Energy Research), Dr. Moksadur Rahman (ABB), Dr. Esin Iplik (Linde) and Prof. Rebei bel Fdhila (Hitachi-Energy Research), for their time, effort, and the arrangement of the great keynote lectures as well as the industrial panel session on Simulation and Modelling in Energy Transition and Future Directions.

In addition, I would like to thank all the authors for once again making these proceedings interesting, diverse, and impressive. Many thanks go to Prof. Esko Juuso (University of Oulu) for managing the conference website. Finally, my sincere gratitude goes to the Linköping University Electronic Press for publishing these proceedings.

I wish you all a joyful time while exploring the papers in this volume!

Prof Konstantinos G. Kyprianidis
Chair of SIMS2023

About SIMS

SIMS is the Scandinavian Simulation Society with members from the five Nordic countries Denmark, Finland, Norway, Sweden and Iceland. The SIMS history dates back to 1959. SIMS practical matters are taken care of by the SIMS board. It consists of two representatives from Denmark, Finland, Norway, Sweden, and Iceland. The SIMS annual meeting takes place at the annual SIMS conference or in connection to international simulation conferences arranged in the Nordic countries.

SIMS webpage: https://scansims.org/

Published by: Linköping University Electronic Press 2023
Linköping Electronic Conference Proceedings, No 200
ISBN 978-91-8075-348-7 (PDF)
ISSN 1650-3686, eISSN 1650-3740

https://doi.org/10.3384/ecp200

SIMS2023 Hosting and Sponsoring Institutions

MDU Mälardalen University
LiU Linköping University
AR Automation Region
EH Electrification Hub
TFEK Tekniska Föreningen i Västerås med Elektriska Klubben

SIMS2023 Conference Chair

Prof. Konstantinos Kyprianidis (Mälardalen University)

SIMS2023 Organizing Committee

Prof. Konstantinos Kyprianidis (Mälardalen University)
Prof. Erik Dahlquist (Mälardalen University)
Dr. Ioanna Aslanidou (Mälardalen University)
Dr. Avinash Renuke (Mälardalen University)
Dr. Hao Chen (Mälardalen University)
Dr. Amare Desalegn Fentaye (Mälardalen University)
Dr. Valentin Scheiff (Mälardalen University)
Dr. Valentina Zaccaria (Mälardalen University)
Dr. Amir Vadiee (Mälardalen University)
Dr. Stavros Vouros (Mälardalen University)
Dr. Xiaodan Shi (Mälardalen University)
Prof. Mirko Morini (University of Parma)
Dr. Gaurav Mirlekar (University of South-Eastern Norway)
Dr. Vadim Engelson (Wolfram MathCore AB)
Prof. Tiina Komulainen (Oslo Metropolitan University)
Prof. Esko Juuso (University of Oulu)
Prof. Lars Eriksson (Linköping University)

SIMS2023 Industrial Organizing Committee

Dr. Lokman Hosain (Hitachi-Energy Research)
Dr. Moksadur Rahman (ABB)
Dr. Esin Iplik (Linde)
Prof. Rebei bel Fdhila (Hitachi-Energy Research)

List of papers

Building heat demand characteristics in a planned city district with low-temperature district heating supply

Karin Israelsson, Vartan Ahrens Kayayan, Fatemeh Johari, Mattias Gustafsson, Magnus Åberg **1**

Heat Demand Modelling for a Sustainable Urban Development Project: A Case Study of Kopparlunden in Västerås, Sweden

Alaa Krayem, Mohammed Guezgouz, Fredrik Wallin **8**

Response Surface Modelling to Reduce CO2 Capture Solvent Cost by Conversion of OZD to MEA

Federico Mereu, Jayangi D. Wagaarachchige, Zulkifli Idris, Klaus-Joachim Jens and Maths Halstensen **14**

Pumped Thermal Energy Storage for Multi-Energy Systems Optimization

Alessandra Ghilardi, Guido Francesco Frate, Antonio Piazzi, Mauro Tucci, Konstantinos Kyprianidis and Lorenzo Ferrari **21**

Applied Machine Learning for Short-Term Electric Load Forecasting in Cities - A Case Study of Eskilstuna, Sweden

Pontus Netzell, Hussain Kazmi and Konstantinos Kyprianidis **29**

Economic investigation of heat pumps for heat recovery from data center

Vahid Zangeneh and Lars Erik Øi **39**

Design of Machine Learning method for decision-making support and reliability improvement in the investment casting process

Antonia Antoniadou, Konstantinos Kyprianidis, Ioanna Aslanidou, Anestis Kalfas and Dimitrios Siafakas **46**

Process Simulation, Dimensioning and Automated Cost Optimization of CO2 Capture

Lars Erik Øi, Shirvan Shirdel, Sumudu Karunarathne and Solomon Aromada **54**

The Effect of Climate and Orientation on the Energy Performance of a Prefab House in Norway

Amirhossein Ghazi, Zahir Barahmand and Lars Erik Øi **62**

Approaching simulation-based controller design: heat exchanger case study

Matias Waller and Leonardo Espinosa Leal **70**

Multimodal sensor suite for identification of flow regimes and estimation of phase fractions and velocities – Machine Learning Algorithms in Multiphase flow metering and Control

Noorain Syed Kazmi, Ru Yan, Håkon Viumdal and Saba Mylvaganam **78**

Simulation of Oil Recovery Through Advanced Wells Using a Transient Fully Coupled Well-Reservoir Model

Madhawee Anuththara, Ali Moradi, Amaranath S. Kumara and Britt M. E. Moldestad **86**

Visualization of Industrial Production Processes using 3D Simulation Software for Enhanced Decision-Making

Akshay Goyal **94**

Optimal indoor temperature flexibility for thermal peak shaving in buildings connected to the district heating network

Mathilda Cederbladh, August Dahlberg, Stavros Vouros, Konstantinos Kyprianidis, Costanza Saletti and Mirko Morini **103**

Sustainability analysis and simulation of a Polymer Electrolyte Membrane (PEM) electrolyser for green hydrogen production

Jordi Béjar Rabascall and Gaurav Mirlekar **110**

Process Simulation and Cost Estimation of CO2 Capture configurations in Aspen HYSYS

Lars Erik Øi, Madhawee Anuththara, Shahin Haji Kermani, Mostafa Mirzapour, Soudeh Shamsiri and Sumudu Karunarathne **118**

Performance assessment of a photovoltaic/thermal-powered absorption chiller for a restaurant

Nima Monghasemi, Stavros Vouros, Konstantinos Kyprianidis and Amir Vadiee **124**

In-Depth System-Level Energy Analysis of Hybrid Electrified Commuter Aircraft for Improved Energy Efficiency

Dimitra Eirini Diamantidou, Valentina Zaccaria and Anestis Kalfas **136**

Phase Fractions and Velocities in Multiphase Flow – Estimation using Sensor Data Fusion and Machine Learning

Andreas Lund Rasmussen, Kjetil Fjalestad, Ru Yan, Håkon Viumdal, Saba Mylvaganam and Tonni Franke Johansen **144**

Process Simulation and Cost Optimization of a Gas based Power Plant including amine based CO2 Capture

Lars Erik Øi and Esmaeil Aboukazempour Amiri **152**

ESP Lifted Oil Field: Core Model, and Comparison of Simulation Tools

Bernt Lie **159**

A Python-based code for modeling the thermodynamics of the vapor compression cycle applied to residential heat pumps

Rebecca Allen, Eirik Svortevik and Henrik Bergersen **167**

Implementation of a bolted joint model in Modelica

Nils Dressler and Lars Eriksson **175**

Numerical Investigation on Performance of Gas Turbine Blade: Effects of simulation Models and Blade Geometry

Heng Hu, Narmin Hushmandi and Magnus Genrup **184**

Insight into the thermodynamic model for reforming of methane over nickel catalyst

Rakhi, Binod R. Giri, Vivien Günther and Fabian Mauss **192**

Developing a Multi-Building Scale Energy Model for a University Campus using URBANopt

Hamed Mohseni Pahlavan and Natasa Nord **198**

Modeling and control of WRRF biogas production

Tiina Komulainen, Bilal Mukhtar, Truls Ødegaard, Hilde Johansen, Kristine Haualand, Kjell Rune Jonassen and Simen Antonsen **206**

A Comparison of Strain Gauge Measurements and FEA for a Confined Channel Geometry Subjected to a Hydrogen-Air Mixture Explosion

Daniel Eckhoff, Magne Bratland and Mads Mowinckel **214**

A Deep Learning Approach for Fault Diagnosis of Hydrogen Fueled Micro Gas Turbines

Muhammad Baqir Hashmi, Mohammad Mansouri, Amare Desalegn Fentaye and Shazaib Ahsan **220**

Hydrodynamic study of a CO2 desorption column using computational fluid dynamics

Sumudu Karunarathne, Kristoffer Eikeseth and Lars Erik Øi **231**

Information extraction from operator interface images using computer vision and machine learning

Eirik Illing, Nils-Olav Skeie and Ole Magnus Brastein **238**

Modeling and identification of the Quanser Aero using a detailed description of friction and centripetal forces

Mathias Dyvik, Didrik Efjestad Fjereide and Damiano Rotondo **246**

Dynamic Modelling and Part-Load Behavior of a Brayton Heat Pump

Matteo Pettinari, Guido Francesco Frate, Konstantinos Kyprianidis and Lorenzo Ferrari. **254**

Future Potential Impact of Wind Energy in Sweden's bidding area SE3

Justin Warners, Stavros Vouros, Konstantinos Kyprianidis, Benders and Nienhuis **262**

Simulation of blue hydrogen production by natural gas in the North Sea

Chidapha Deeraksa, Britt Margrethe Emilie Moldestad and Nora Cecilie Ivarsdatter S. Furuvik **271**

The Impact of Autonomous Inflow Control Valve on Enhanced Oil Recovery in SAGD Application

Farhan Hasin Alam, Amin Tahami, Nora C.I. Furuvik, Britt M.E. Moldestad and Soheila Taghavi **278**

Estimation of effluent nutrients in municipal MBBR process

Tiina Komulainen, Abdul Malik Baqeri, Einar Nermo, Arvind Keprate, Torgeir Saltnes, Katrine Marsten Jansen and Olga Korostynska **286**

Banks of Gaussian Process Sensor Models for Fault Detection in Wastewater Treatment Processes

Heidi Lynn Ivan and Jean-Paul André Ivan **294**

Automatic Translator from System Dynamics to Modelica with Application to Socio-Bio-Physical Systems

John Tinnerholm, Mariano Zapatero, Adrian Pop, Peter Fritzson and Rodrigo Castro **302**

Data-driven reinforcement learning-based parametrization of a thermal model in induction traction motors

Anas Fattouh and Smrutirekha Sahoob **310**

Simulation of distribution system for low temperature district heating in future urban areas – Case study of a planned city district in Gävle

Oskar Olsson, Mattias Gustafsson and Magnus Åberg **318**

Models for a hydropower plant: a review

Tajana Nepal, Diwakar Bista, Thomas Øyvang and Roshan Sharma **326**

Traceable System of Systems Explorations Using RCE Workflows

Jorge Lovaco, Ingo Staack and Petter Krus **339**

Thermodynamics analysis of a novel compressed air energy storage (CAES) system combined with SOFC-MGT and using low grade waste heat as heat source

Chen Yang and Li Sun **347**

Green production of dimethyl ether (DME) - indirect conversion of synthesis gas produced from biomass

Marianne Eikeland, Sebastian Larsen, Oliver Numme, Eivind Johan Trasti and Terje Bråthen **358**

Enhancing Indoor Environmental Simulations: A Comprehensive Review of CFD Methods

Shahrzad Marashian, Amir Vadiee, Omid Abouali and Sasan Sadrizadeh **365**

Mapping Simulation optimization requirements for construction sites: A study in heavy-duty vehicles industry

Abdulkarim Habbab, Anas Fattouh, Bobbie Frank, Koteshwar Chirumalla and Markus Bohlin **372**

An embedded industrial control framework for model predictive control of district heat substation

Joakim Örneskans, Konstantinos Kyprianidis, Stavros Vouros and Gunnar Bengtsson **379**

Development of a MATLAB-based code for quantification of effective void space in porous pavement

Rebecca Allen, Berthe Dongmo-Engeland and Saja Al-Batat **386**

Machine learning assisted adaptive heat load consumption forecasting in district heating network

Avinash Renuke, Stavros Vouros and Konstantinos Kyprianidis **393**

Retrofitting Biomass Combined Heat and Power Plant for Biofuel Production

Hao Chen, Daheem Mehmood, Erik Dahlquist and Konstantinos Kyprianidis **405**

Building heat demand characteristics in a planned city district with low-temperature district heating supply

Israelsson, Karin[a,*] Ahrens Kayayan, Vartan[b] Johari, Fatemeh[a] Gustafsson, Mattias[b] Åberg, Magnus[a]

[a]*Division of Civil Engineering and Built Environment, Uppsala University* [b]*Faculty of Engineering and Sustainable Development, Gävle University*

[*]karin.israelsson123@gmail.com

Abstract

Due to desirable emission reductions and population growth, increasing energy demand is identified as a dire issue for energy systems. The introduction of low-energy building districts enables increased energy system efficiency. This study's aim is twofold. Firstly, an extensive urban building energy model is used to simulate the hourly use and geographic distribution of the heat demand for residential and commercial buildings that are to be supplied by a low-temperature district heating system. The simulated buildings are a part of a planned city district, located in Gävle, Sweden. Two building energy performance cases are studied; one where all buildings are assumed to be of Passive House standard, and one where the building energy performance is in line with conventional new-building regulations in Sweden. Secondly, one specific building is modeled in detail and simulated in the building energy simulation software IDA ICE to investigate what building heating system is best suited for low-temperature heat supply. The temperature demands of floor heating and ventilation with heat recovery are investigated and compared to conventional water-based radiators. The building's temperature demand results can be used when designing a low-tempered district heating system which will provide the supply temperature to identify a compatible heating system technique. Varying supply temperature demand will enable optimization for choosing building heating systems and consequently, possible cost reductions. The results could be used as an example for future city district planning as well as presenting relevant heating systems for low-temperature district heating.

1 Introduction

Energy demand is increasing at a problematic rate for regional and global systems due to increasing population and the desire to reduce greenhouse gas (GHG) emission (Energimyndigheten, 2022a). The Swedish housing sector utilized roughly 34% of the total energy use in Sweden in the year 2022, where the majority of energy was used for space heating (SH) and domestic hot water (DHW) (Boverket, 2023). With an increasing construction rate, the energy systems tend to be further burdened with higher energy demands (Naturvårdsverket, 2022). Studies, such as the one carried out by (Abu Bakar et al., 2015), have found that it is possible to reduce energy systems by implementing energy-efficient buildings. District heating (DH) is the most common heat source (>50%) for residential and commercial buildings in Sweden (Energimyndigheten, 2022b). To comply with the EU's climate goal of lowered energy use, studies have been conducted to explore the possibility of implementing low-temperature district heating systems, where the supply and return temperatures are lowered (Lund et al., 2014). Low-temperature district heating has two distinct advantages: the ability to incorporate lower temperature sources that would otherwise be wasted and the reduction of distribution losses. These in turn can lead to a reduction in GHG emissions. This would imply lower energy use as well as distribution losses and higher demand on the building envelope. Johari et al. (2023) and Reinhart & Cerezo Davila (2016), among others, developed and investigated urban building energy models (UBEMs) in the interest of facilitating planning of city districts. To enable the implementation of low-temperature DH, further studies have been conducted to investigate suitable heating systems for buildings. Hasan et al. (2009) found low-temperature water heating systems, radiators, and floor heating, to be applicable to such a heat source while maintaining comfortable indoor air temperature.

The work in this paper is part of a larger work where the aim is to use the 3D plans for a coming city district and simulate the energy use on a building level. Then simulate losses and mass flows in different low-temperature DH-systems and finally, simulate supply and return temperatures in a building with different heating systems using water as distribution media. This paper focuses on simulating the heating demand on a building level from 3D plans of an upcoming city district and simulating different heating systems inside a building connected to a DH system. A planned city district located in the northern parts of Gävle, Sweden, will be used as a case study to inves-

tigate the importance of building energy performance and heating system temperature demands for applying low-temperature DH. The first will be investigated by simulating varying building energy performances to later be used for planning a DH. The second will simulate inlet temperatures to space heaters to determine the best-suited technique for low-temperature DH supply.

2 Background

The planned city district (Näringen) is located north of Gävle, Sweden, and will be used as a reference case for this study. The district intends to be one of Europe's most sustainable city districts with roughly 6000 residential apartments and 450000 m^2 commercial space (Gävle kommun, 2021). The buildings are to be supplied with low-temperature DH, further addressed as 4th generation DH (4GDH), generally defined by supply and return temperatures of $55/30°C$ (Lund et al., 2014). 4GDH has been proposed as a method to reduce energy losses and incorporate waste heat to decarbonize heating needs (Connolly et al., 2014). One advantage is that once established there is flexibility to incorporate carbon-neutral thermal energy sources. This paper focuses on the requirements of the shell and indoor heat delivery, i.e. radiators, required for 4GDH. District heating is considered a favorable heat source when an area's heat density is at least $40-50\ kWh/m^2$ (Frederiksen & Werner, 2014). Below this threshold, thermal losses in the pipes leading to and within the area are too high. Lower energy demands can also be uneconomic to justify the infrastructure investment of district heating. A shell with a higher U-value leads to lower energy demand but may push the heat density below the recommended for district heating network integration. Both heat density and heat delivery inside the apartments are identified as research gaps in the field of 4GDH.

3 Methodology

To investigate the influence of energy performance of buildings and their space heating systems, in particular, water radiators, on low-temperature DH, two studies were made. In the first study, an UBEM method developed by Johari et al. (2023), was used for simulating the energy performance of buildings based on the criteria for Passive House standard and the latest Swedish building codes with and without heat recovery ventilation system. Second, to find the best-suited system for low-temperature DH supply, the inlet water temperature to conventional radiators, low-temperature radiators, and floor heating was determined using IDA ICE (AB, 2023).

As plans for the new city district are not yet final, assumptions were made when simulating in IDA ICE and UBEM. Information regarding the geographical position, number of floors, and building type was retrieved from a project description (Gävle kommun, 2021). Figure 1 shows the planned city district's buildings as well as the development phases 1-11.

Figure 1. Geographical positioning of the planned city district including development phases 1-11.

3.1 Weather data

Due to buildings' energy use's dependency on weather, both models utilize data for a typical meteorological year (TMY) for Gävle from the PVGIS Online Tool (European Commission, n.d.). The measured ambient temperature for 2022 was used when simulating the annual heat demand to use data consistent with the current temperature profile instead of the historical one. The data was collected from Gävle Energi's database, which is the company responsible for managing the DH system in Gävle. In IDA ICE, the wind profile was set to represent urban conditions.

3.2 Estimation of heat demand in the planned city district

To estimate the energy use of a large set of buildings located in the planned city district, a bottom-up physics-based UBEM developed in Johari et al. (2023) was used to estimate space heating (SH) and domestic hot water (DHW) use. The model was originally made for residential buildings. However, for the scope of this study, it was further extended to cover non-residential buildings, i.e., administrative and office buildings, as well. The key difference between residential and non-residential buildings was assumed to be centered around occupancy and load. Therefore, a new occupancy profile was attributed to non-residential buildings. Using the methodology suggested in Sandels et al. (2015), the occupancy profile for non-residential buildings was developed from a stochastic model for occupants' presence, use of electrical appliances, and lighting. As for the use of domestic hot water, it was set to zero (Sveby, 2010).

In this study, two types of building standards were used for modeling and simulation of buildings. The

first was the latest Swedish building codes and minimum requirements for new constructions (in short, referred to as BBR) (Boverket, 2020). The second was the Passive House standard (in short, PH), which was proven to result in higher building energy performance than BBR. In the UBEM, the thermal properties of these standards were set according to the presented data in Table 1. In addition, for the case of PH, a mechanical heat recovery ventilation system with a minimum effect of 75 % was also considered in the model(FEBY, 2018). The window-to-wall ratio was assumed to be 20 % (Cerezo et al., 2017).

Table 1. U-values (W/m^{2}K) for PH and BBR.

Building	U_{wall}	U_{floor}	U_{roof}	U_{window}
PH	0.10	0.09	0.09	0.8
BBR	0.17	0.15	0.12	1.2

3.2.1 UBEM simulation scenarios

The importance of building energy performance when computing heat demand was simulated in three scenarios. First, all buildings were simulated based on BBR without ventilation heat recovery. Second, the energy performance was increased by implementing ventilation heat recovery. Third, all buildings were simulated as PH with ventilation heat recovery.

3.3 Estimation of the temperature demand

Urban scale models are rather simplified and therefore unable to represent a detailed low-temperature system (Johari et al., 2023). Therefore, it was chosen to use the building simulation tool IDA ICE to investigate the best-suited heating system for low-temperature DH supply. The supply temperature of water-based high-temperature radiators, low-temperature radiators, and floor heating were studied respectively. One three-story residential building with simple geometry was modeled in detail to represent BBR-building with ventilation heat recovery, $\eta_v = 75\%$. The floor area was estimated in the geographic information software QGIS to be 423.6 m^2 with interior and exterior ceiling heights at 2.5 and 3 m, respectively. A study by (Johari et al., 2022) has shown that simplified one-building modeling in IDA ICE results in limited errors. Hence, the building was assumed one zone per floor and evenly distributed windows with a window-to-wall ratio 20 % (Figure 2).

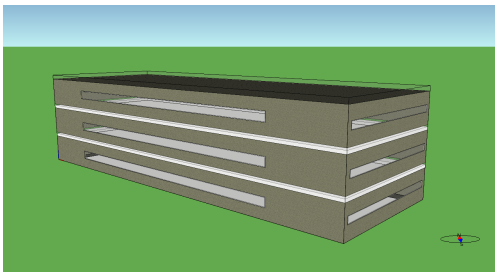


Figure 2. Building model in IDA ICE.

According to BBR the air change was set to 0.35 L/s·m^2 and DHW use 25 kWh/$m^2 \cdot A_{temp} \cdot y$ (Boverket, 2020). The average living space per person is 42 m^2/person in Sweden which was used to estimate occupancy of 10.09 people per floor with absence between 7 am to 5 pm and 80 W heat emission. This was further used to estimate heat emission from lighting and appliances 932.7 W/floor (Sveby, 2009).

A study by Hasan et al. (2009) was used to simulate low- and high-temperature profiles of the water supply (Figure 3). The study used design temperatures 21°C and -26°C for summer and winter, respectively, resembling the Swedish climate.

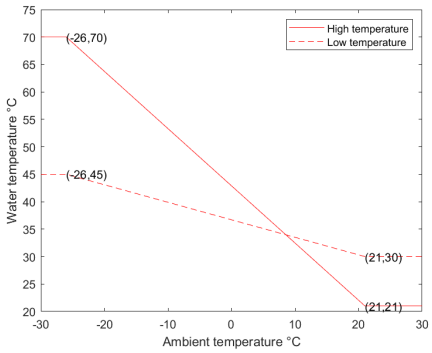


Figure 3. Water supply low- and high-temperature profiles to building heating system.

3.3.1 IDA ICE simulation scenarios

A 100 W/m^2 floor heated area radiator was placed under the window on the building's long side on each floor, generating a heat emissive area of 11.7 m^2. The design indoor temperature at maximum power was set to 21°C and ΔT at 10°C (Hasan et al., 2009). With the PI-regulator, IDA ICE estimated the water mass flow to be 1.0 kg/s. The supply water temperatures were then varied between high and low to simulate conventional high-temperature radiators and low-temperature radiators. Neither radiator system is supplemented with comfort floor heating. The supply temperature is determined by the outdoor temperature, and the mass flow varies depending on the heating demand.

35 circuits/floor at 12 m^2, placed 3 cm into the floor material, was used to simulate floor heating. Each circuit emitted 50 W/m^2 and had a design ΔT at 10°C as in Hasan et al. (2009). With temperature control and PI-regulator, a constant water mass flow of 0.014 kg/s was estimated in IDA ICE. The floor heating system was considered a low-temperature profile as in Figure 3. Three independent scenarios are tested:

- High-temperature radiators only
- Low-temperature radiators only
- Floor heating system only

4 Results

In this section, the results from the implementation of the methodology are presented in two parts. First, it is shown that PH has the lowest heat demand but an effective ventilation heat recovery has a large impact on the total heat demand. Implementation of varying building energy performances could be more beneficial to make DH possible, which in turn has a positive impact on the energy system. Floor heating showed to need the lowest supply temperatures of the investigated heating systems.

4.1 Area heat demand and heat density

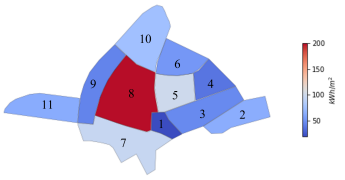
Results show that the total heat demand (DHW and SH) for the simulated area decreases from 130, 85, and 79 GWh/year with increasing building energy performance. Figure 4 shows the annual heat demand for each building of the district. A minor decrease (approximately 6%) in heat demand is seen when simulating buildings with PH standard compared to BBR-building when both include ventilation heat recovery. Heat recovery decreases the heat demand by approximately 34% when simulated on BBR-buildings. Thus, effective heat recovery has been shown to have a larger impact than materials with low U-values on a city district's heat demand. Further, Figure 4, also illustrates the buildings with the highest heat demand. This is due to its floor area and wall ratio and consequently inadequate solar heating. The building with the highest yearly heat demand decreases from 2980, 1930 to 1880 MWh for BBR-building without ventilation, BBR-building with ventilation, and PH.

Figure 5 shows each development phase's heat density. Phase 8 is well above the requirement (40 kWh/m^2) for DH for each simulated energy performance. Whereas phases 1 and 4 are always below this value. This result might be used to optimize the district for DH use.

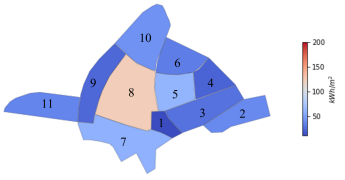
(a) BBR without ventilation heat recovery.

(b) PH with ventilation heat recovery.

Figure 4. Distributed total heat demand (MWh) for PH (with ventilation heat recovery) and BBR (without heat recovery).



(a) BBR-building without ventilation heat recovery.



(b) PH with ventilation heat recovery.

Figure 5. Heat density (kWh/m^2) for each development phase.

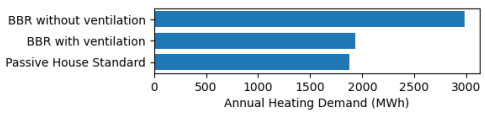


Figure 6. Results for the annual energy demand (MWh).

4.2 Individual building temperature demand

Figure 7 shows the temperature demand of the simulated space heaters and the hours the heaters are in use, i.e. the hours with heat demand in descending energy demand. Inlet and outlet temperature is set to the indoor temperature ($21°C$) when the heater power is $< 1\ W$. Floor heating is shown to require the lowest temperatures, maximum inlet/outlet temperature at $27/24°C$. The results showed a small ΔT of $1-3°C$. This is beneficial for maintaining comfort and even heat emissions. The two types of radiators require maximum inlet temperature of $41°C$ and $58°C$, low temperature and conventional, respectively. Almost identical outlet temperatures were simulated, following the indoor temperature, indicating effective heat emission regardless of the space heater.

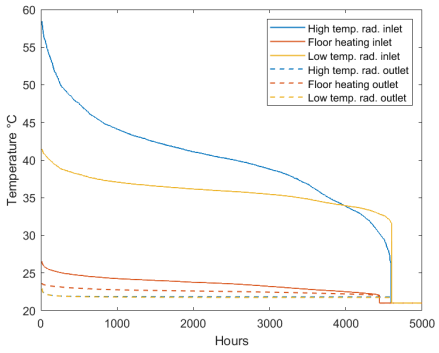


Figure 7. In- and outlet temperature to the heating system in BBR-building for winter.

Figure 7 shows that low-temperature radiators have a steady temperature curve to maintain the desired indoor temperature. However, high-temperature radiators enable fluctuations in the outlet temperature making it possible to use lower temperatures when possible as well as raising the heat when needed. After roughly 4000 h can the high-temperature radiators use lower temperatures than low-temperature radiators and still maintain an indoor temperature of $21°C$. Although counterintuitive, it reflects the assumption that the low-temperature radiator will require higher inlet temperatures at outdoor temperatures above $10°C$ which could reflect comfort floor heating in the cited paper (Hasan et al., 2009). At approximately 4500 h there is no longer a heat demand and the space heaters are therefore shut off. The principal results are however that the system is feasible to provide comfort even at lower inlet temperatures.

5 Discussion

The results from Figure 4 further the discussion and cost calculations on ventilation heat recovery versus low U-value constructions. As the results show heat recovery has a larger impact on reduction in heat demand of approximately 34% than possibly expensive PH construction of 6% compared to BBR-building without heat recovery. These gains need to be compared with the cost for each adjustment. A life cycle assessment (LCA) should also be conducted to understand the environmental impact. According to the current plans, a few residential buildings use a lot of energy when the majority have equivalent heat demand. These buildings should be evaluated and redrawn to e.g., minimize transmission losses through the building envelope and maximize solar gains.

When planning a city district supplied with DH, Figure 5 can be used to optimize the heat demand by varying building energy performance for different areas aiming at the threshold value at 40 kWh/m^2. By altering building energy performance in the different phases, the heat density can be made better suited for DH, allowing higher utilization of waste heat from e.g. industries. The result may also be used to identify issues in the current development plans such as the number of stories.

The low level of detail in buildings, both when simulating in IDA ICE as well as UBEM, results in simplified but sufficient calculations (Johari et al., 2022). Floor heating is shown to be the space heater best suited for low-temperature DH with a maximum temperature demand of $27°C$ and the lowest return temperature. When other aspects such as the initial installation cost of the heating system and slower response time to indoor or outdoor temperature changes are taken into account, other technologies can be favorable. One example can be the low-temperature radiator system that has a maximum supply temperature demand of $41°C$ which is still below the definition for 4GDH ($55/30°C$).

6 Conclusion

Improving efficiency can reduce the thermal energy demand of buildings but for DH systems to remain viable changes to heat delivery need to be made. To answer the questions of what impact a building's energy performance has on a city district's heat demand and which type of space heater is best suited for low-temperature DH, simulations in an UBEM and IDA ICE were conducted. Results showed PH has the lowest heat demand of 79 GWh/y compared to the reference case with BBR-building without ventilation heat recovery of 130 GWh/y. However, ventilation heat recovery seems to have a larger impact on a building's heat demand (34%) than construction with lower U-values (39%) compared to the reference case. Floor heating is the space heater best suited for low-temperature DH with a maximum inlet temperature of $27°C$ and the lowest return temperature of the investi-

gated heating systems. Low-temperature radiators are also a good fit with 4GDH but demand a higher temperature of 41°*C* rather than floor heating's 27°*C*.

Acknowledgement

This work was supported by the Swedish Energy Agency, project "Samspel mellan värme, kyla och elanvändning i ett bostadsområde med en hög andel solcellsproducerad el" [grant number P2022-00442].

References

AB, E. S. (2023). *Ida indoor climate and energy.* Retrieved from equa.se/en/ida-ice

Abu Bakar, N. N., Hassan, M. Y., Abdullah, H., Rahman, H. A., Abdullah, M. P., Hussin, F., & Bandi, M. (2015). Energy efficiency index as an indicator for measuring building energy performance: A review. *Renewable and Sustainable Energy Reviews*, *44*, 1-11. Retrieved from https://www.sciencedirect.com/science/article/pii/S1364032114010703 doi: https://doi.org/10.1016/j.rser.2014.12.018

Boverket. (2020). *Bfs 2020:4 - bbr 29.* Retrieved from https://rinfo.boverket.se/BFS2011-6/pdf/BFS2020-4.pdf

Boverket. (2023). *Bygg- och fastighetssektorns energianvändning uppdelat på förnybar energi, fossil energi och kärnkraft.* Retrieved from https://www.boverket.se/sv/byggande/hallbart-byggande-och-forvaltning/miljoindikatorer---aktuell-status/energianvandning/

Cerezo, C., Sokol, J., AlKhaled, S., Reinhart, C., Al-Mumin, A., & Hajiah, A. (2017). Comparison of four building archetype characterization methods in urban building energy modeling (ubem): A residential case study in kuwait city. *Energy and Buildings*, *154*, 321-334. Retrieved from https://www.sciencedirect.com/science/article/pii/S0378778817314743 doi: https://doi.org/10.1016/j.enbuild.2017.08.029

Connolly, D., Lund, H., Mathiesen, B., Werner, S., Möller, B., Persson, U., … Nielsen, S. (2014). Heat roadmap europe: Combining district heating with heat savings to decarbonise the eu energy system. *Energy Policy*, *65*, 475-489. Retrieved from https://doi.org/10.1016/j.enpol.2013.10.035 doi: https://doi.org/10.1016/j.enpol.2013.10.035

Energimyndigheten. (2022a). *Energiläget 2022.*

Energimyndigheten. (2022b). *Fjärrvärme.* Retrieved from https://www.energimyndigheten.se/trygg-energiforsorjning/el/trygg-fjarrvarme/

European Commission. (n.d.). *Photovoltaic geographical information system.* Retrieved from https://re.jrc.ec.europa.eu/pvg_tools/en/#TMY

FEBY. (2018). *Feby18.* Retrieved from https://www.feby.se/

Frederiksen, S., & Werner, S. (2014). *Fjärrvärme och fjärrkyla* (1:4 ed.). Lund: Studentlitteratur AB.

Gävle kommun. (2021). *Förstudie stadsomvandling näringen.* Retrieved from https://www.gavle.se/kommunens-service/bygga-trafik-och-miljo/planer-och-samhallsbyggnadsprojekt-i-gavle/pagaende-byggprojekt-i-gavle/naringen/

Hasan, A., Kurnitski, J., & Jokiranta, K. (2009). A combined low temperature water heating system consisting of radiators and floor heating. *Energy and Buildings*, *41*(5), 470-479. Retrieved from https://www.sciencedirect.com/science/article/pii/S0378778808002570 doi: https://doi.org/10.1016/j.enbuild.2008.11.016

Johari, F., Munkhammar, J., Shadram, F., & Widén, J. (2022). Evaluation of simplified building energy models for urban-scale energy analysis of buildings. *Building and Environment*, *211*(108684). Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-427499 doi: 10.1016/j.buildenv.2021.108684

Johari, F., Shadram, F., & Widén, J. (2023). *Urban building energy modeling from geo-referenced energy performance certificate data: Development, calibration, and validation.*

Lund, H., Werner, S., Wiltshire, R., Svendsen, S., Thorsen, J. E., Hvelplund, F., & Mathiesen, B. V. (2014). 4th generation district heating (4gdh) integrating smart thermal grids into future sustainable energy systems. *Energy (Oxford)*, *68*, 1-11.

Naturvårdsverket. (2022). *Klimatet och bygg- och fastighetssektorn.* Retrieved from https://www.naturvardsverket.se/amnesomraden/klimatomstallningen/omraden/klimatet-och-bygg--och-fastighetssektorn/

Reinhart, C. F., & Cerezo Davila, C. (2016). Urban building energy modeling - a review of a nascent field. *Building and Environment*, *97*, 196-202. Retrieved from https://www.sciencedirect.com/science/

article/pii/S0360132315003248 doi: https://doi.org/10.1016/j.buildenv.2015.12.001

Sandels, C., Widén, J., & Nordström, L. (2015). Simulating occupancy in office buildings with non-homogeneous markov chains for demand response analysis. In *2015 ieee power energy society general meeting* (p. 1-5). doi: 10.1109/PESGM.2015.7285865

Sveby. (2009). *Brukarindata för energiberäkningar i bostäder*. Retrieved from https://www.sveby.org/wp-content/uploads/2011/06/brukarindata_bostader.pdf

Sveby. (2010). *Brukarindata för energiberäkningar i kontor - vägledning*. Retrieved from https://www.sveby.org/wp-content/uploads/2012/01/Brukarindata-Kontor.pdf

Heat Demand Modelling for a Sustainable Urban Development Project: A Case Study of Kopparlunden in Västerås, Sweden.

Alaa Krayem[a], Mohammed Guezgouz [a,*], Fredrik Wallin [a]

Future Energy Center, Department of Energy, Building and Environment, Mälardalen University, Västerås, Sweden

mohammed.guezgouz@mdu.se

Abstract

As cities grow and develop, urban planners face an increasing challenge to create more sustainable and environment friendly communities. The Kopparlunden district in Västerås, Sweden, is no exception, with plans underway to transition the area to a more sustainable neighborhood. To assist this effort, this paper presents a simple grey box modeling approach to predict the heat demand of eight buildings in the area. As the city transforms from a historical industrial district to a mixed district with residential buildings, shops, and offices, the model will allow urban planners to predict their new heat demand. The model is calibrated using a genetic algorithm, then validated using real historical data. The results show a good accuracy of the model and highlight the importance of increasing the insulation efficiency of the walls in the modelled buildings. The model can be used to predict the heat demand variations, with minimum error of 2.49 kW and up to 16.6 kW for large buildings. The model highlights the importance of energy modeling for urban development projects and shows its significance as a tool to aid in decision-making towards sustainable and more efficient urban areas.

1. Introduction

With the increased growth of urban population and urban energy use, cities around the world are facing an increasing challenge to provide sustainable and energy efficient environment for their residents. The building sector is one of the major contributors of CO2 emissions worldwide, and thus presents a great potential to reduce energy use and associated emissions. Urban planners and policymakers are to plan and implement large scale improvements of buildings' energy performance, which imposes significant challenges and issues related to the complexity and scale of the urban environment (Hong et al., 2020; Keirstead et al., 2012), policy and regulatory frameworks essential to adequately support and /or incentivize sustainable and energy efficient practices (Economidou et al., 2020; Strielkowski et al., 2019), funding and securing financial resources for projects implementation (Alam et al., 2019; Bertoldi et al., 2021; Sebi et al., 2019), data availability and accessibility for informed decision making, and long term planning and adaptation with consideration of future needs and changing circumstances. In this context, the European Commision recently revised the Energy Performance of Buildings Directive (EPBD) under the "Fit for 55" package (Wilson, 2022), and introduced stricter regulations. The revised EPBD aims at accelerating the renovation rates, targets the 15% of EU buildings that perform the worst, and establishes high energy performance standards. Notably, every building should achieve at least a Class E on the revised A-G energy performance scale by 2030.

The Kopparlunden district in Västerås, Sweden, is no exception to this global and regional trend. As part of a larger effort to foster sustainability, plans are underway to transition the district into a more sustainable neighborhood.

Building Energy Modelling (BEM) became an indispensable tool for building professionals and energy policy makers to optimize the design, operation, and energy efficiency of buildings (Al-Homoud, 2001; Reinhart & Cerezo Davila, 2016). BEM can be performed at the individual building level, up to the urban level (Urban Building Energy Modelling – UBEM). Its approaches comprise three main categories: white-box models, black-box models and grey box models (Foucquier et al., 2013). The white-box models are based on physical equations that describe the underlying mechanisms of the building. They offer transparency and understanding of the physical phenomena involved, allowing for accurate predictions and optimization, as well interpretability of the results. However, there are drawbacks to consider, such as the complexity of dealing with complex systems, and the time-consuming nature of model development (Harish & Kumar, 2016). Black-box models on the other hand, are purely data driven models. They use actual data and perform statistical analysis to capture the correlation between the building energy use and

operation data (Li & Wen, 2014). Grey-box models represent a hybrid approach that combines physical and empirical equations to achieve a close approximation of the underlying physical representation (Harb et al., 2016). They are utilized when there are partial information or incomplete data, allowing for flexibility and adaptation in handling discrepancies, and providing a more robust modeling framework (Zhao & Magoulès, 2012).

Discrepancies between a model's predictions and actual energy use are inevitable. To reduce the entailing mismatch, calibration is applied. It is a process of changing and fine-tuning the model's parameters and input assumptions to guarantee that the simulated energy performance matches the actual energy use of the building (Chong et al., 2021). It consists of comparing the model's predictions to measured data in the building and making modifications to increase the model's accuracy and credibility. The calibration of BEM can be either manual, where it relies on the modeler expertise, or automated, where an objective function is set to match the simulation results with the measured data (Coakley et al., 2014; Hou et al., 2021). Among the popular calibration techniques are optimization evolutionary algorithms, such as genetic algorithms (Lara et al., 2017).

In this paper, we present a study that focuses on the simulation of eight buildings in the Kopparlunden area of Västerås, Sweden. Our main objective is to develop a grey box model capable of predicting the hourly heat demand for each building under steady state conditions. Despite the simplicity of the model, we ensured its accuracy through a careful calibration process using a genetic algorithm. By incorporating this calibration technique, we fine-tune the model's parameters to improve its performance and align it with measured data. The resultant model achieves a good balance between simplicity and accuracy, making it a useful and effective tool for predicting heat demand in the investigated buildings. Our findings demonstrate the successful use of a basic yet calibrated grey box model, emphasizing its utility in giving vital insights for energy efficiency and decision-making in building energy management.

2. Methodology

2.1. Case study

Kopparlunden, an industrial area in Västerås dating back to 1898, holds historical significance. Situated in close proximity to the city center, as depicted in Figure 1 on the map, the majority of its buildings still retain their original character and were originally utilized for metal industry purposes. However, the evolving landscape has seen a shift in usage, with the buildings now serving as offices or stores, accommodating nearly 200 companies in the vicinity. Recognizing the potential for optimizing the local area, plans have been set in motion to revitalize Kopparlunden into a contemporary residential space, integrating modern housing, commercial establishments, and workspaces[1].

The municipality is dedicated to maximizing the energy efficiency of the area and has actively collaborated with various partners to oversee the implementation of the plan. The transformation of Kopparlunden is part of a multi-step strategy that the municipality and building companies have collectively committed to. While many aspects of the project, such as the size and functionality of the buildings, have been determined, finer details regarding the architectural design and specific shape are still under consideration. However, at this stage, it is possible to make preliminary assessments of certain parameters, such as the current heat demand, which is the primary focus of the current study.

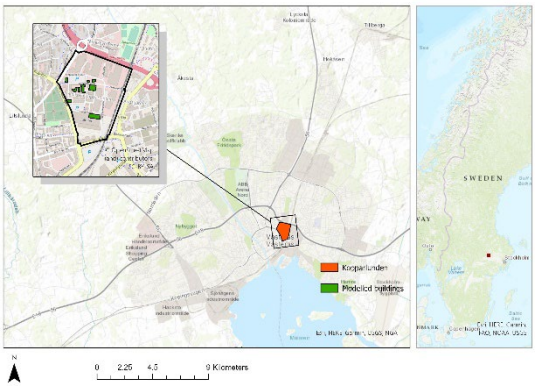


Figure 1: Kopparlunden district, in close proximity to Västerås center. The modelled buildings are shown in green.

2.2 Data Collection

The buildings simulated in this study are highlighted in Figure 2. Buildings data were obtained from the NRGYHUB dataset (Krayem et al., 2021). The data included buildings' perimeters, areas, and heights. The real heat demand data was obtained from Mälarenergi at hourly level for the year 2019. The outside temperature was downloaded from ERA5[2] for the same year at the hourly level.

[1] https://www.archus.se/kopparlunden-fran-ett-stangt-industriomrade-till-en-levandestadsdel/ (accessed 26/6/2023)

[2] https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=form (accessed 26/6/2023)

2.3. Model and assumptions

It is assumed that the building heat balance is given using the following equation:

$$P_D + P_{people} = P_s \quad (1)$$

where P_s, power loss by transmission; P_D, generated heat power (only district heating and generated heat from the occupants). To simplify the model, losses by ventilation, unintended ventilation and air leakage were neglected.

For each element of the building, the transmission loss is calculated by the following equation:

$$P_s = U\ .\ A\ .\ (T_{in} - T_{out}) \quad (2)$$

where U, heat transfer coefficient in W/m^2.°C, A, area in m^2, T_{in}, indoor temperature in °C, and T_{out}, outdoor temperature in °C. The total transmission loss of the building is the sum of the individual transmission loss of each element.

The U_{value} of different elements of old buildings were assumed from the literature (Liu et al., 2014). The U_{value} of the floor was assumed 0.22W/m^2.°C and that of windows 1.3 W/m^2.°C. The walls assembly, with an overall thickness of 0.4m, were considered to be a composite structure comprising, in sequence, brick, concrete, wood, insulation material, wood, and a final concrete layer. The respective thicknesses of these materials are 0.09m for concrete, 0.06m for timber, and 0.1m for insulation. Corresponding thermal conductivity values for these materials are delineated in Table 1. The overall U_{total} of the walls is equal to the reciprocal of its total resistance R_{total}, which is calculated as follows:

$$R_{total} = \sum \frac{d_i}{U_i} \quad (3)$$

where d_i represents the thickness of layer i and U_i its corresponding thermal conductivity. The areas of the walls, floors and ceilings were estimated from the shapefiles. The area of the windows, which were assumed double glazed, was then calculated using the window to wall ratio from Table 1.

For T_{in} , it is assumed to be 21°C to ensure indoor comfort.

The internal generated heat power is considered mainly generated from occupancy and is calculated using the following equation:

$$P_{people} = A_{floor}.n_{floors}.P_{ph}.r_p \quad (3)$$

where A_{floor}, a building's floor area, n_{floor}, number of floors, P_{ph}, heat generated per person, assumed to be 80W/person, and r_p, the person ratio in 1/m^2 and it is assumed to be one person per 35m2. The values are based on the Swedish National Board of Housing, Building, and Planning (BBR)[3].

Table 1: Range of values of inputs estimated with the genetic algorithm.

Variables	Range of variations
Window to wall ratio	0.1 – 0.65
Concrete heat conductivity	1.3 – 2
Wood heat conductivity	0.12 – 0.16
Insulation heat conductivity	0.06 – 0.1
Wall resistance indoor	0.1 – 0.16
Wall resistance outdoor	0.02 – 0.06

The walls of all buildings are assumed to be composed of double layers of concrete and wood with an insulation layer in between. Several key inputs related to the materials properties and buildings construction (shown in Table 1) remain indeterminate due to unavailable or ambiguous data. To address these uncertainties and ensure the reliability of the model, a calibration process was conducted using a genetic algorithm (Martínez et al., 2020). The objective of the calibration (cost function) was to minimize the Root Mean Square Error (RMSE) between the simulated heat demand and the observed heat demand data. This iterative process involved fine-tuning some of the model's parameters to achieve a higher level of accuracy in predicting the heat demand. During the calibration process, multiple design variables were considered, as outlined in Table 1. These inputs played a crucial role in optimizing the model's performance and aligning it with the actual heat demand patterns recorded in the Kopparlunden area. Careful selection and adjustment of these variables contributed to improving the model's capability to simulate the complex heat demand patterns observed in the buildings. The variations' range of the concrete heat conductivity was obtained from (Misri et al., 2018) and for wood from (Pásztory et al., 2020).

By iteratively adjusting and refining these design variables, the aim was to enhance the model's accuracy and its ability to capture the variation nature and seasonal patterns of the heat demand profiles of different buildings.

3. Results

Among the buildings studied, Building II exhibited the highest level of accuracy in terms of heat demand prediction, with an RMSE of approximately 2 kW, as shown in Table 2. Conversely, the first building demonstrated the largest deviation from the actual heat demand, resulting in an RMSE of 16 kW. This discrepancy can be attributed to several factors,

[3] Boverkets föreskrifter om ändring av verkets föreskrifter och allmänna råd (2016:12) om fastställande av byggnadens energianvändning vid normalt brukande och ett normalår, BFS 2017:6

Table 2: The estimated inputs of the model after calibration using the genetic algorithm.

Building number	I	II	III	IV	V	VI	VII	VIII
Window to wall ratio	0.10	0.14	0.10	0.10	0.10	0.10	0.10	0.10
U_{value} of window	0.85	1.28	0.85	0.85	0.92	0.85	0.85	0.85
Concrete heat conductivity	1.30	1.78	1.30	1.30	1.35	1.30	1.30	1.37
Wood heat conductivity	0.12	0.14	0.12	0.12	0.13	0.12	0.12	0.14
Insulation heat conductivity	0.06	0.07	0.06	0.06	0.07	0.06	0.06	0.08
Wall resistance indoor	0.16	0.16	0.16	0.16	0.12	0.16	0.16	0.15
Wall resistance outdoor	0.06	0.05	0.06	0.06	0.04	0.06	0.06	0.03
RMSE [kW]	16.60	2.49	5.62	10.92	14.71	4.93	8.51	4.97

including the lack of detailed information regarding the construction materials utilized in Building 1. Additionally, the accuracy of the geometry data employed in the model significantly influences the predictive performance.

Figure 2 (scatter plot) illustrates a comparison between the actual heat demand points and the corresponding predicted values. The closeness of the points to the diagonal red line indicates the degree of agreement between the actual and predicted values. A strong correlation is observed for most of the considered buildings, as signifies the tight clustering of points around the diagonal red line, while deviations suggest a divergence between the actual and predicted values.

The analysis reveals that buildings 1, 5, and 7 exhibit the highest discrepancies between predicted and actual values. Remarkably, these buildings are also the largest in the study, boasting significant annual heat demands of 153.52 MWh, 349.95 MWh, and 189.84 MWh, respectively. This highlights a notable limitation in the model's accuracy when estimating heat demand for sizable and complex buildings. Achieving more accurate results in these cases necessitates a more detailed approach that considers additional factors.

Figure 3 illustrates the distribution of losses among the buildings based on the optimal design variables obtained from the calibration process. It is evident that wall losses contribute the most significant proportion of total losses for all buildings, closely followed by losses through the ceiling. Conversely, losses through the windows are relatively low due to the smaller window-to-wall ratio considered and the utilization of effective insulation materials.

The utilization of a grey-box model in this study provides a straightforward approach to estimate building heat demand. However, it is important to recognize that higher levels of accuracy may necessitate a substantial amount of data. A larger and more detailed dataset would have contributed to enhancing the precision of the model's predictions. While detailed data might be available for specific or individual projects, and it is possible to achieve detailed data collection, it might not always feasible for large scale modeling, such as in Urban Building Energy Modeling (Hao & Hong, 2021; Wong et al., 2021), given the vast heterogeneity in buildings and associated operational variables. Different modeling approaches are adopted, ranging from physics-based to statistical-based methods (Swan & Ugursal, 2009). Each comes with its own set of advantages and limitations, depending on the availability of data and the specific objectives of the analysis.

The heat losses shown in Figure 3 highlights the importance of insulation to reduce the wall heat losses for buildings. The findings suggest that improving the insulation and thermal characteristics of the walls could lead to substantial reductions in energy losses. Furthermore, using highly thermal resistant materials in the ceiling can also contribute to minimizing overall heat losses.

By focusing on these key areas of concern, such as wall and ceiling insulation, building operators and policymakers can effectively enhance energy efficiency and reduce heating demands. This understanding of the relative contributions of different building components to heat losses offers valuable insights for implementing targeted interventions and developing sustainable heating practices in the Kopparlunden area.

4. Summary and Discussions

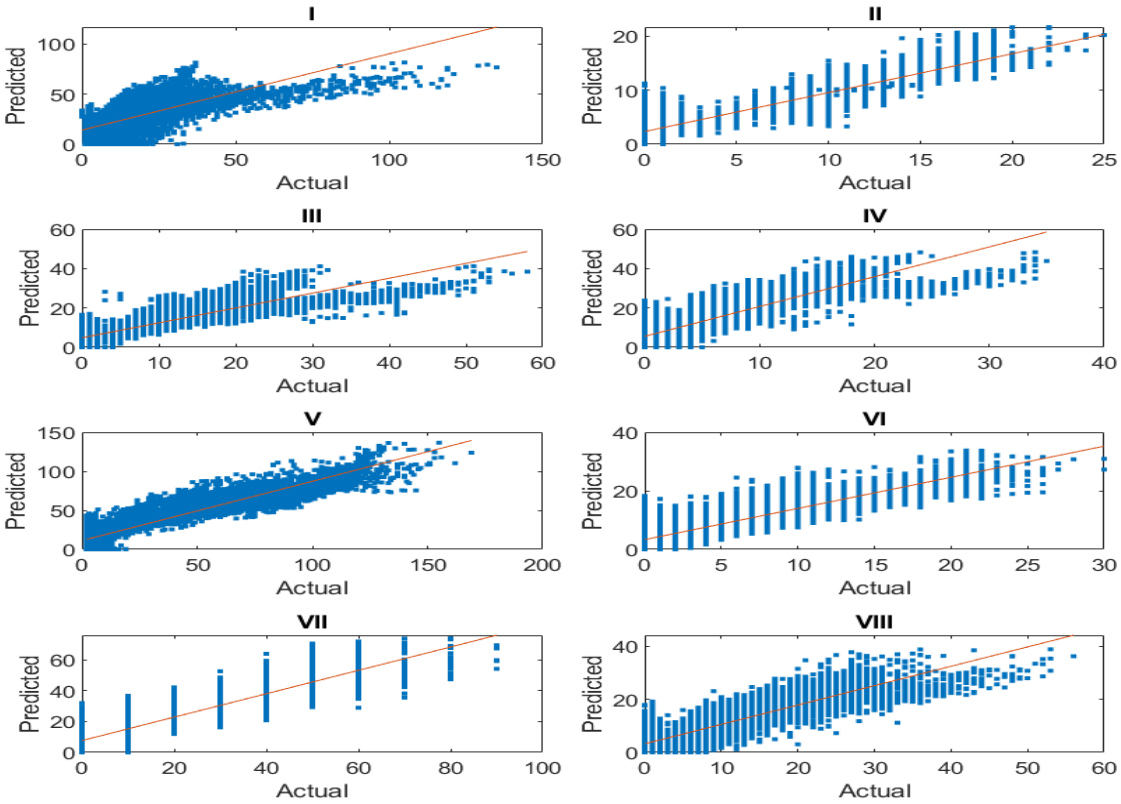


Figure 2: Comparison between the simulated and actual heat demand per building.

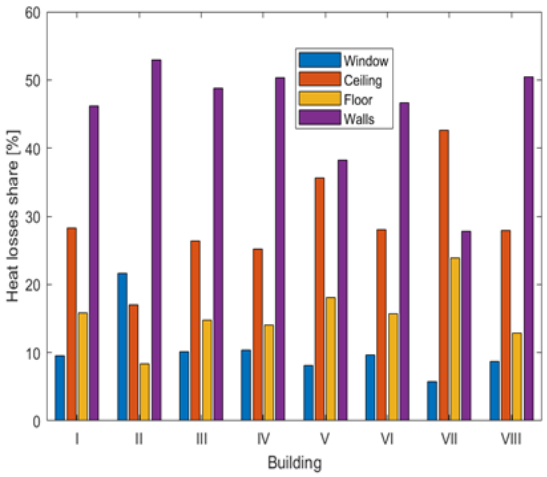


Figure 3: Estimated heat losses per building element, for the eight buildings modelled in this study.

While the results obtained from the model offer valuable insights into the fluctuations of heat demand in the studied buildings, it is crucial to realize the inherent limitations associated with this approach. Grey-box models may rely on simplified assumptions and estimated parameters and may not fully capture the intricate complexities of real-world systems. Nevertheless, this methodology serves as a valuable tool for providing initial estimations of heat demand and can serve as a starting point for further analysis and refinement.

By acknowledging both the strengths and limitations of the grey-box model and considering the availability and quality of data, researchers and practitioners can make informed decisions regarding energy management and optimization strategies. Future actions should concentrate on enhancing the model's accuracy through the incorporation of more detailed information, improved geometry data, and potential exploration of alternative modeling techniques to achieve even higher levels of predictive performance.

Acknowledgment

We acknowledge support from the KK-foundation (Grant #20200073) through the project FLEXERGY, and from the Reduction and Reuse of energy with interconnected Distribution and Demand (R2D2) project.

References

Alam, M., Zou, P. X. W., Stewart, R. A., Bertone, E., Sahin, O., Buntine, C., & Marshall, C. (2019). Government championed strategies to overcome the barriers to public building energy efficiency retrofit projects. *Sustainable Cities and Society*, *44*, 56–69. https://doi.org/https://doi.org/10.1016/j.scs.2018.09.022

Al-Homoud, M. S. (2001). Computer-aided building energy analysis techniques. *Building and Environment*, *36*(4), 421–433. https://doi.org/https://doi.org/10.1016/S0360-1323(00)00026-3

Bertoldi, P., Economidou, M., Palermo, V., Boza-Kiss, B., & Todeschi, V. (2021). How to finance energy renovation of residential buildings: Review of current and emerging financing instruments in the EU. *WIREs Energy and Environment*, *10*(1). https://doi.org/10.1002/wene.384

Chong, A., Gu, Y., & Jia, H. (2021). Calibrating building energy simulation models: A review of the basics to guide future work. *Energy and Buildings*, *253*(October). https://doi.org/10.1016/j.enbuild.2021.111533

Coakley, D., Raftery, P., & Keane, M. (2014). A review of methods to match building energy simulation models to measured data. In *Renewable and Sustainable Energy Reviews* (Vol. 37, pp. 123–141). https://doi.org/10.1016/j.rser.2014.05.007

Economidou, M., Todeschi, V., Bertoldi, P., D'Agostino, D., Zangheri, P., & Castellazzi, L. (2020). Review of 50 years of EU energy efficiency policies for buildings. *Energy and Buildings*, *225*, 110322. https://doi.org/https://doi.org/10.1016/j.enbuild.2020.110322

Foucquier, A., Robert, S., Suard, F., Stéphan, L., & Jay, A. (2013). State of the art in building modelling and energy performances prediction: A review. *Renewable and Sustainable Energy Reviews*, *23*, 272–288. https://doi.org/10.1016/j.rser.2013.03.004

Harb, H., Boyanov, N., Hernandez, L., Streblow, R., & Müller, D. (2016). Development and validation of grey-box models for forecasting the thermal response of occupied buildings. *Energy and Buildings*, *117*, 199–207. https://doi.org/https://doi.org/10.1016/j.enbuild.2016.02.021

Harish, V. S. K. V. S. K. V, & Kumar, A. (2016). A review on modeling and simulation of building energy systems. *Renewable and Sustainable Energy Reviews*, *56*, 1272–1292. https://doi.org/10.1016/j.rser.2015.12.040

Hong, T., Chen, Y., Luo, X., Luo, N., & Lee, S. H. (2020). Ten questions on urban building energy modeling. *Building and Environment*, *168*, 106508. https://doi.org/10.1016/j.buildenv.2019.106508

Hou, D., Hassan, I. G., & Wang, L. (2021). Review on building energy model calibration by Bayesian inference. *Renewable and Sustainable Energy Reviews*, *143*, 110930. https://doi.org/https://doi.org/10.1016/j.rser.2021.110930

Keirstead, J., Jennings, M., & Sivakumar, A. (2012). A review of urban energy system models: Approaches, challenges and opportunities. *Renewable and Sustainable Energy Reviews*, *16*(6), 3847–3866. https://doi.org/10.1016/j.rser.2012.02.047

Krayem, A., Wallin, F., & Yan, J. (2021). Open Urban Data Portal for Collaborative Research and Innovation. *Energy Proceedings*. https://doi.org/10.46855/energy-proceedings-8378

Lara, R. A., Naboni, E., Pernigotto, G., Cappelletti, F., Zhang, Y., Barzon, F., Gasparella, A., & Romagnoni, P. (2017). Optimization Tools for Building Energy Model Calibration. *Energy Procedia*, *111*, 1060–1069. https://doi.org/https://doi.org/10.1016/j.egypro.2017.03.269

Li, X., & Wen, J. (2014). Review of building energy modeling for control and operation. In *Renewable and Sustainable Energy Reviews* (Vol. 37, pp. 517–537). https://doi.org/10.1016/j.rser.2014.05.056

Liu, L., Moshfegh, B., Akander, J., & Cehlin, M. (2014). Comprehensive investigation on energy retrofits in eleven multi-family buildings in Sweden. *Energy and Buildings*, *84*, 704–715. https://doi.org/https://doi.org/10.1016/j.enbuild.2014.08.044

Martínez, S., Eguía, P., Granada, E., Moazami, A., & Hamdy, M. (2020). A performance comparison of multi-objective optimization-based approaches for calibrating white-box building energy models. *Energy and Buildings*, *216*, 109942. https://doi.org/10.1016/J.ENBUILD.2020.109942

Misri, Z., Ibrahim, M. H. W., Awal, A. S. M. A., Desa, M. S. M., & Ghadzali, N. S. (2018). Review on factors influencing thermal conductivity of concrete incorporating various type of waste materials. *IOP Conference Series: Earth and Environmental Science*, *140*, 012141. https://doi.org/10.1088/1755-1315/140/1/012141

Pásztory, Z., Fehér, S., & Börcsök, Z. (2020). The effect of heat treatment on thermal conductivity of paulownia wood. *European Journal of Wood and Wood Products*, *78*(1), 205–207. https://doi.org/10.1007/s00107-019-01470-3

Reinhart, C. F., & Cerezo Davila, C. (2016). Urban building energy modeling – A review of a nascent field. *Building and Environment*, *97*, 196–202. https://doi.org/10.1016/j.buildenv.2015.12.001

Sebi, C., Nadel, S., Schlomann, B., & Steinbach, J. (2019). Policy strategies for achieving large long-term savings from retrofitting existing buildings. *Energy Efficiency*, *12*(1), 89–105. https://doi.org/10.1007/s12053-018-9661-5

Strielkowski, W., Volkova, E., Pushkareva, L., & Streimikiene, D. (2019). Innovative Policies for Energy Efficiency and the Use of Renewables in Households. *Energies*, *12*(7). https://doi.org/10.3390/en12071392

Swan, L. G., & Ugursal, V. I. (2009). Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable and Sustainable Energy Reviews*, *13*(8), 1819–1835. https://doi.org/https://doi.org/10.1016/j.rser.2008.09.033

Wilson, A. (2022). *Revision of the Energy Performance of Buildings Directive: Fit for 55 package*.

Zhao, H. X., & Magoulès, F. (2012). A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews*, *16*(6), 3586–3592. https://doi.org/10.1016/j.rser.2012.02.049

Response Surface Modelling to Reduce CO_2 Capture Solvent Cost by Conversion of OZD to MEA

Federico Mereu [a], Jayangi D. Wagaarachchige [a], Maths Halstensen [a], Zulkifli Idris [a], Klaus-J. Jens [a, *]

[a] *University of South – Eastern Norway, Kjølnes ring 56, 3918 Porsgrunn, Norway*
Klaus.J.Jens@usn.no

Abstract

The increasing CO_2 concentration in the atmosphere is the most urgent global challenge. The most mature CO_2 abatement option is post-combustion CO_2 capture employing Monoethanolamine (MEA) solvent. One challenge of using MEA is its in-service degradation to 2-oxazolidinone (OZD), a heterocyclic five-membered organic ring compound. Furthermore, OZD degrades more MEA leading to CO_2 capture solvent loss and hence increased operational cost. It is therefore of interest to investigate methods to convert OZD back to MEA. This work reports the conversion of 2-oxazolidinone to MEA by heat treatment at an alkaline condition. Raman spectroscopy and Ion-Exchange chromatography were applied to qualify and quantify the reaction. The optimal reaction parameters were identified by an experimental design model using the Response Surface Methodology (RSM). A second-order model with three variables and five levels of focus was employed, with the OZD conversion percentage as the response. This methodology was chosen because such a model could estimate the main effects, interactions and quadratic terms by relying on a relatively small number of experiments. 17 experimental runs were designed by the software using this method. At a reaction time of 35 minutes, reaction temperature of 100°C, and 2.5 mole of hydroxide per mole of OZD resulted in a complete conversion of OZD to MEA.

1. Introduction

Carbon dioxide capture and storage (CCS) is so far considered the most promising technology to sequestrate CO_2 from large emission point sources (Rochelle, 2009). Post-combustion carbon capture (PCC) gas-liquid chemical adsorption is the predominant CCS technology today because of the development status (*US Department of Energy*, 2017) and that it can be retrofitted to existing CO_2 emitting plants.

Aqueous 30 wt% Monoethanolamine (MEA) is one of the most investigated CO_2 absorption solvents due to its good operational properties and relatively low price (Kohl & Nielsen, 1997; Buvik, 2021). Figure 1 shows the typical CO_2 absorption-desorption process scheme. The flue gas enters the absorber bottom after pre-treatment and flows upwards while the solvent solution e.g., aqueous 30 wt% MEA moves downwards in counter flow. Through a contact of these two streams, the CO_2 flue gas content is absorbed into the amine solution, forming mainly amine carbamate (equation (1)) which can release CO_2 upon heating to 120-140°C in the process stripping section according to equation (2) (Eimer, 2014). For simplicity, MEA is expressed by R-NH_2, where R stands for a -CH_2-CH_2OH group.

$$2R-NH_2 + CO_2 \rightarrow R-NH_3^+ + R-NH-COO^- \quad (1)$$

$$R-NH_3^+ + R-NH-COO^- \xrightarrow{Heat} 2R-NH_2 + CO_2 \quad (2)$$

However, aqueous MEA solvent has a high energy need in the solvent regeneration section and it is degrading in service due to contact with air in the absorption section and high temperature in the stripping section of the process (Fredriksen & Jens, 2013). These degradation reactions reduce solvent absorption capacity. Furthermore, these degradation products have to be removed and replaced with fresh MEA solvent which adds to operational cost. This degradation (reclaimer waste) varies in the range of 0.1-14.9 kg waste /ton CO_2 captured (IEAGHG, 2014).

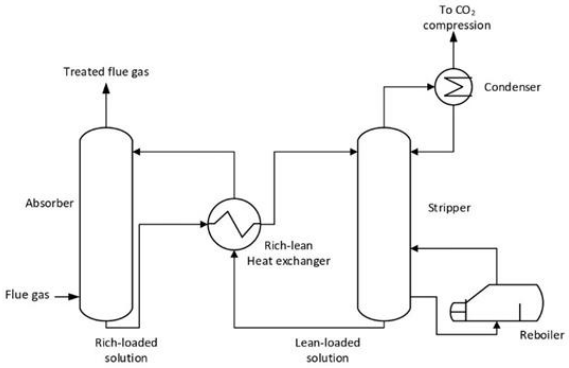


Figure 1: Basic schematic diagram of the chemical absorption-desorption CO_2 capture process.

The first step of thermal MEA degradation is the formation of oxazolidone (Davis & Rochelle, 2009; Dyen & Swern, 1967; Poldermann et al., 1955) from the reaction of MEA with CO_2. Several patents (Miller, 1985; Pottiez & Verbeest, 1972; Snoble, 1981; Turoff et al., 2008) claim hydrolytic alkaline splitting of the alkanolamine derived oxazolidone back to the original alkanolamine and carbonate anion as shown in Figure 2.

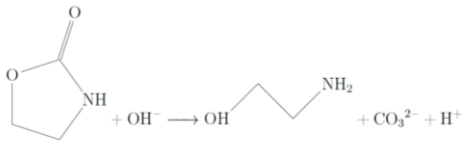


Figure 2: Reaction for Splitting of oxazolidone by alkali to MEA and a carbonate anion.

It is therefore desirable to understand the optimum reaction conditions for splitting of oxazolidone type thermal degradation products back to the original CO_2 capturing alkanolamine. Hence, this work determines optimal reaction conditions for splitting of oxazolidone by Design of Experiment utilizing the Response Surface Methodology (RSM) (Myers et al., 2016). Furthermore, a process integration concept into a CO_2 capture plant is proposed.

2. Methodology

2.1 Design of Experiments

A screening design (Eriksson et al., 2008; Esbensen & Swarbrick, 2017) was used to identify the most important parameters, thus reducing the number of experiments needed. For the optimization experiments, a Central Composite Design (Eriksson et al., 2008; Esbensen & Swarbrick, 2017) was used to vary the parameters at 5 levels as indicated in Figure 3.

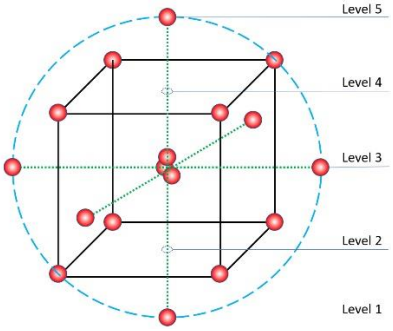


Figure 3. Central composite design with three center samples resulting in 17 experiments. The 5 different levels of the vertical variable are indicated.

The optimal reaction parameters were identified by an experimental design model using the RSM provided by the JMP software.

2.2 Chemicals

Chemicals were used as received and are summarized in Table 1. All aqueous solutions were prepared with Milli-Q® water (18.2 MΩ·cm at 25˚C).

Table 1. Chemicals used in the experiments.

Chemical name	**CAS number**	**Supplier**	**Mole Fraction purity (%)**
Ethanolamine	141-43-5	EMSURE	≥ 99.5
2-Oxazolidinone	497-25-6	Sigma-Aldrich	98
Sodium carbonate	497-19-8	Sigma-Aldrich	99.9

2.3. Chemical Analysis

Cation chromatography and Raman spectroscopy were used to determine the reference concentrations.

2.3.1 Cation chromatography

The samples were analyzed by a Dionex 5000 Cation chromatograph controlled by Chromeleon® software and equipped with a Dionex IonPac CS16 2 mm column. The eluent was methanesulfonic acid run at a gradient method (Table 2) and a constant flow rate of 0.5 mL/min at 60˚C temperature. Samples were diluted with 0.3 ppm Li^+ containing Milli-Q water (internal standard) to a factor of 900 and filtered before injection using a 0.2 mm syringe filter.

Table 2. Eluent gradient concentrations

Time [min]	**Concentration [mM]**
0	6
13	8
25	55
30	60
32	60
34	6
40	0

2.3.2 Raman spectroscopy.

The spectra were acquired with a RXN2 Raman spectrometer fitted with a 785nm laser delivered by Kaiser Optical Systems Inc. The samples were placed inside a black sample holder covered with an aluminium foil to suppress background light and were measured using a fiber optic immersion probe. The probe was washed with deionized water

followed by acetone before each measurement. The analysis exposure time was 30 seconds with 6 scans for each sample. A blank water sample preceded each analysis run. All Raman spectra were baseline corrected using the Whittaker filter (Eilers, 2003; Whittaker, 1922) (available in PLS toolbox in the MATLAB ®software suite) before further processing.

2.4 Oxazolidone (or 2-Oxazolidinone) splitting experiment.

The experimental set-up is shown in Figure 4. A sample of aqueous 3M oxazolidone stock solution and NaOH pellets or 0.1M NaOH solution was placed in a flask connected to a reflux condenser and heated under stirring to a pre-determined temperature and time. The pH of the reaction mixture was monitored by a pH electrode.
In a typical experiment, 25g 3M aqueous oxazolidone solution was titrated with 0,1 M NaOH solution using a Mettler Toledo T50 titrator pH 14 or a total solution volume of 80 ml whatever happed first. The latter produced a pH value of 12.4 at 80 ml solution. The solution was heated to 110°C for a few minutes, cooled to room temperature and analyzed for oxazolidone conversion by Raman spectroscopy. An alternative typical experiment of the above procedure substituted the 0.1 M NaOH solution by anhydrous NaOH pellets followed by water dilution to pH14. The aqueous dissolution of NaOH pellets is an exothermic process. Hence, the reaction solution was cooled to room temperature before Raman spectroscopic analysis.

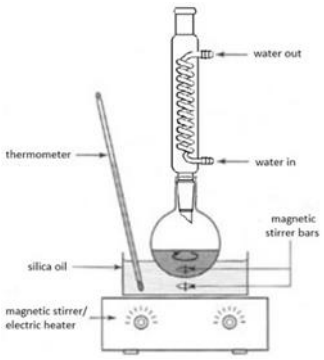
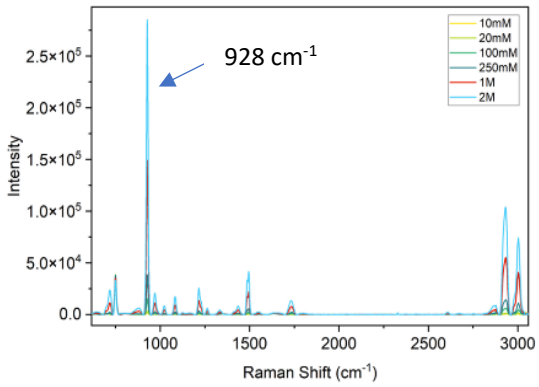


Figure 4: Experimental set-up

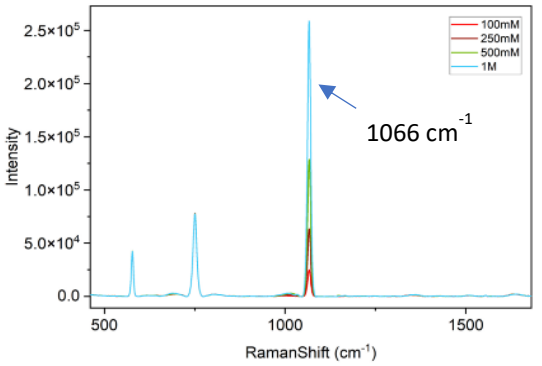
3. Results and Discussions

3.1. Chemical Calibration for MEA and OZD Quantification

Concentrations of OZD were quantified using Raman Spectroscopic analysis. The spectral peaks of oxazolidone at 928 cm^{-1} (Figure 5 (a)) (ref: McDermott (1986)) and of Na2CO3 at 1066 cm^{-1} (Figure 5 (b)) were chosen for subsequent qualitative and quantitative analysis.



(a)



(b)

Figure 5: (a) Raman spectra of oxazolidone (b) sodium carbonate at various concentrations

Cation chromatography was then used for the quantification of MEA. The MEA peak was identified in the chromatogram (Figure 6) by spiking it with an authentic MEA sample. Quantitative analysis was based on a calibration curve in the appropriate concentration range.

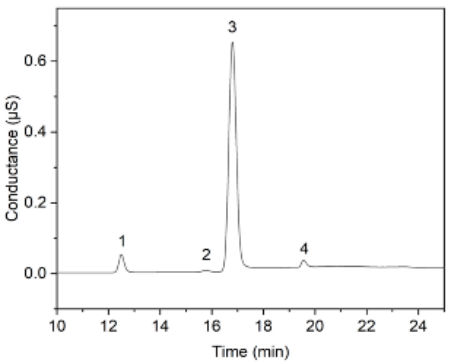


Figure 6: MEA chromatogram: MEA (3); Na^+ (1); $NH4^+$ (2); K^+ (4).

3.2 Optimization of conversion of OZD to MEA

3.2.1 Phase 1

screening investigations. The initial experiments in this phase verified oxazolidone splitting to MEA. As a preliminary study to verify the conversion of OZD to MEA, four tests were carried out using the alternative typical experiment procedure with solution reflux at 120°C and 2 hour duration (Table 3).

Table 3: Parameters for the oxzolidone (OZD) splitting experiment

Experiment	OZD conc. [mM]	OH^- /OZD	Temp. [°C]	Time [min]
3	200	28	90	100
4	400	14	140	60
5	600	12	100	120
6	80	73	130	30

Three parameters were identified based on the initial screening experiments. The parameters were varied at two levels and three parameters were identified. Reaction time, reaction temperature, and the relationship between mole of hydroxide per mole of OZD were all found as the most significant contributors to the conversion of OZD to MEA.

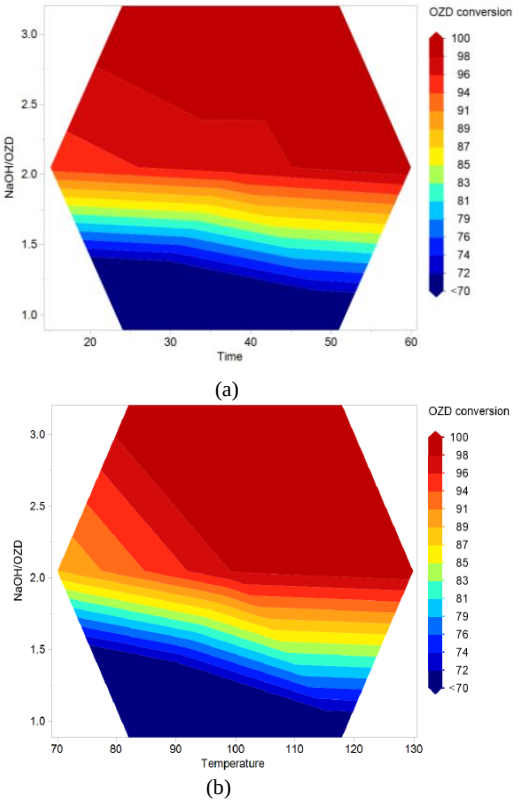
3.2.1 Phase 2

Optimization model for OZD conversion percentage. In phase 1, 3 three variables were chosen for the second phase: temperature (°C), time (min), and OH^-/OZD molar ratio. The Response Surface Methodology was chosen because it can fit a second-order polynomial model that estimates main effects, interactions, and quadratic terms relying on a relatively small number of runs. The quadratic terms are useful because they can capture a possible curvature in the relationship between the response and the experimental factors. The three main variables are represented by the coefficients A, B, and C; the three two-way interactive terms are described by AB, AC, and BC; the three quadratic terms by A^2, B^2, and C^2.

With these considerations, the number of central points (cp) was set to 3, the number of factorial points (Cube) was 8, and the number of axial points was 6. Three additional runs (No NaOH (3a), 1a and 2a) were add to the design resulting in a total of 20 experimental runs. Experimental plan details and the responses (OZD conversion %) are given in Table 4. Figure 7 (a) and (b) show the response surface contour plots of OZD conversion for 100°C (temperature) and 35 min (time), respectively.

Table 4: Optimization model data and results

Exp. No.	Data			Results
	Temp. [ºC]	Time [min]	OH^- /OZD	OZD conversion [%]
1	100	38	4	100
2	118	51	3.2095	99.88
3	118	24	3.2095	99.94
4	82	24	3.2095	99.83
5	82	51	3.2095	99.93
6	70	38	2.05	89.2
7	130	38	2.05	99.53
8	100	38	2.05	98.83
9	100	60	2.05	99.58
10	100	38	2.05	98.92
11	100	15	2.05	94.52
12	100	38	2.05	98.69
13	82	51	0.8905	53.34
14	82	24	0.8905	48.17
15	118	24	0.8905	59.52
16	118	51	0.8905	74.63
17	100	38	0.1	10.48
1a	130	60	0.8905	72.21
2a	130	60	0.1	13.54
3a	130	60	0	0



(a)

(b)

Figure 7: OZD conversion in function of OH-/OZD mole ratio (a) and time at 100°C (fixed temperature) (b) and temperature at 35 min (fixed time)

The summarizing prediction performance (Table 5) and the coefficient values of the optimization model (Table 6) indicate it to be satisfactory.

Table 5: Optimization model: prediction performance

RMSE	**P-value**	**R^2**	**Adj. R^2**
2.1958	<0.0001	0.99727	0.993167

Table 6: F-ratios and P-values of coefficients values (A: temperature, B: time, C: OH^-/OZD)

Coefficient	**F-ratio**	**P-value**
A	37.7291	0.00085
B	12.5534	0.01217
C	803.7899	0.00000
AB	2.4178	0.17096
AC	27.5179	0.00193
BC	10.5144	0.01763
A^2	3.4492	0.11266
B^2	0.1520	0.71013
C^2	507.2047	0.00000

The model shows several sets of possible variable combinations (Table 7) depending on desired reaction condition application.

Table 7: Model responses for selected variable sets

OH^- /OZD	**Temp. [ºC]**	**Time [min]**	**OZD conversion achieved [%]**
1	130	60	83.46
1.5	130	60	98.59
1.8	100	60	96.97
1.8	120	45	99.01
2	100	47	99.12
2	110	36	99.11
2.2	100	24	98.96
2.5	76	15	98.99

Analytic chemical evidence of reaction response for selected variable combinations is given in Figure 8.

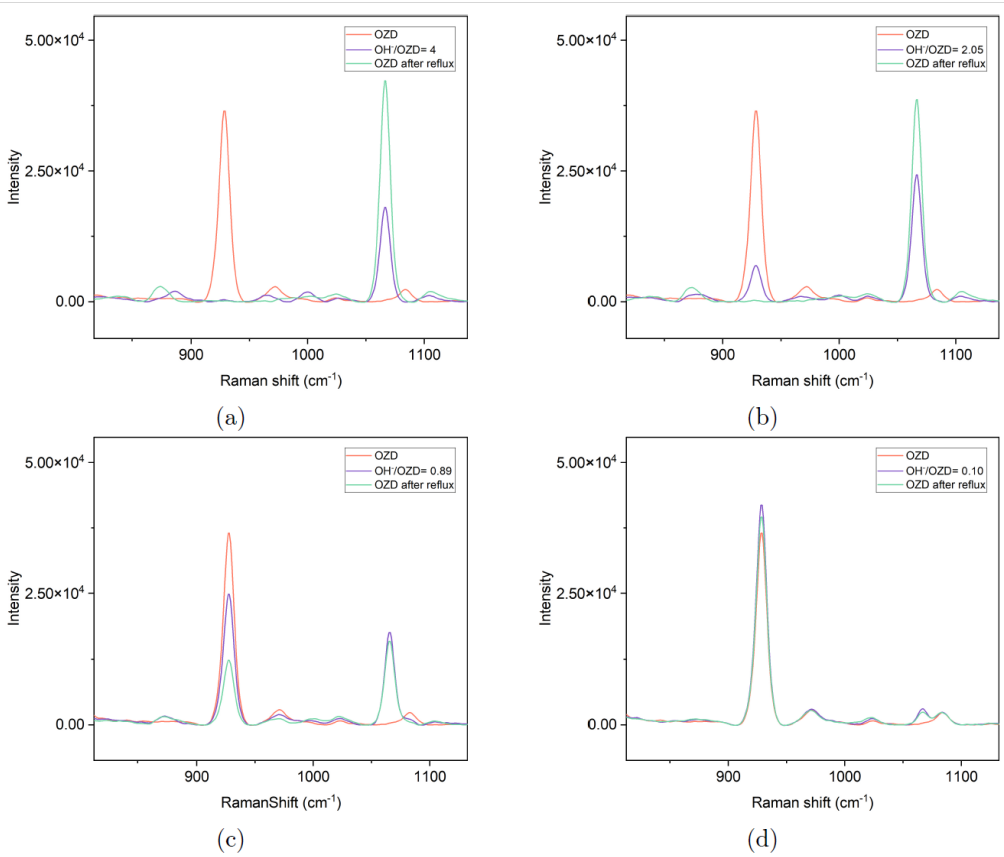


Figure 8. Raman spectra of selected experiments: Axial C (high) (a); Axial B (high) (b) and Axial C (low) (d).

3.3 Potential integration of the oxazolidone splitting reaction into a carbon dioxide capture unit.

The optimal reaction conditions for oxazolidone (OZD) splitting (Table 7) depend mostly on the OH−/OZD mole ratio. The reaction itself could be run in the reboiler of the CO_2 stripper or in a separate reactor connected to it (Figure 1). This would provide a temperature of the reaction vessel in excess of 100ºC (Poldermann et al., 1955). As suggested by a US patent (Turoff et al., 2008), a slipstream of the stripper bottoms could be sent to a separate reactor where NaOH is added, the reaction is conducted and the recovered alkanolamine is separated from the caustic by phase splitting. Some caustic could be recycled to the reaction vessel while the rest is treated as solvent reclaiming waste. The US patent (Turoff et al., 2008) teaches this proposal for splitting of the oxazolidone hydroxypropyloxazolidone (HPOZD) to di-isopropanolamine (DIPA) and CO_2. Our optimal reaction conditions for OZD splitting would fit well with the above process integration proposal.

4. Conclusion

- Society needs to stabilize and reduce CO_2 emissions. Large-scale post-combustion carbon capture is expected to be necessary in the near future. Our contribution puts focus on the cost reduction of aqueous alkanolamine carbon capture solvent.
- Alkanolamine recovery by splitting of oxazolidone (OZD) to the original MEA alkanolamine solvent and CO_2 can be satisfactorily modeled and optimized by the RSM method.
- The model variables are in order of importance OH-/OZD ratio, reaction temperature and reaction time.
- A CO_2 capture process integration concept for splitting of OZD to MEA is proposed.

References

Buvik, V. Høisæter, K. Vevelstad, S. & Knuutila, H. (2021). A review of degradation and emissions in post-combustion CO2 capture pilot plants. *International Journal of Greenhouse Gas Control* 106 (2021) 103246.

Davis, J., & Rochelle, G. (2009). Thermal degradation of monoethanolamine at stripper conditions. *Energy Procedia*, *1*(1), 327-333. https://doi.org/https://doi.org/10.1016/j.egypro.2009.01.045

Dyen, M. E., & Swern, D. (1967). 2-oxazolidones. *Chem Rev*, *67*(2), 197-246. https://doi.org/10.1021/cr60246a003

Eilers, P. H. C. (2003). A Perfect Smoother. *Analytical Chemistry*,*75*(14), 3631-3636. https://doi.org/10.1021/ac034173t

Eimer, D. (2014). *Gas treating : absorption theory and practice*. John Wiley & Sons, Inc.

Eriksson, L., Johansson, E., Kettaneh-Wold, N., Wikström, C., & Wold, S. (2008). *Design of experiments : principles and applications* (3 ed.). Umetrics, Umeå.

Esbensen, K. H., & Swarbrick, B. (2017). *Multivariate Data Analysis: An introduction to Multivariate Analysis, Process Analytical Technology and Quality by Design* (6th ed.). CAMO software AS.

Fredriksen, S. B., & Jens, K.-J. (2013). Oxidative Degradation of Aqueous Amine Solutions of MEA, AMP, MDEA, Pz: A Review. *Energy Procedia*, *37*, 1770-1777. https://doi.org/https://doi.org/10.1016/j.egypro.2013.06.053

IEAGHG, (2014). Evaluation of Reclaimer Sludge Disposal from Post-Combustion CO2 Capture. (March 2014).

Kohl, A., Nielsen, R. (1997). *Gas Purification*. Gulf Prof. Publ.

Miller, A. E. (1985). *Catalytic process for converting 2-oxazolidinones to their corresponding alkanolamines*.

Myers, R. H., Montgomery, D. C., & Anderson-Cook, C. M. (2016). *Response surface methodology: process and product optimization using designed experiments*. John Wiley & Sons.

Poldermann, L. D., Dillon, C. P., & Steele, A. B. (1955). Why MEA Solution Breaks Down in Gas Treating Service. *Oil and Gas Journal*, 4.

Pottiez, F., & Verbeest, R. (1972). Process for the regeneration of alkanolamines by distillation, alkali hydroxide treatment, heat, phase separation and distillation. In: Google Patents.

Rochelle, G. T. (2009). Amine Scrubbing for CO_2 Capture. *Science*, *325*(5948), 1652-1654. https://doi.org/10.1126/science.1176731

Snoble, K. A. J. (1981). Catalytic process for converting oxazolidinones to their corresponding aminoalcohols. In: Google Patents.

Turoff, M. L. H., Cummings, A. L., Waite, S. W., & Horan, R. L. (2008). Process for improving the conversion of oxazolidones to alkanolamines. In: Google Patents.

US Department of Energy (2017) Accelerating Breakthrough Innovation in Carbon Capture, Utilization and Storage.

Whittaker, E. T. (1922). On a new method of graduation. *Proceedings of the Edinburgh Mathematical Society*, *41*, 63-75.

Pumped Thermal Energy Storage for Multi-Energy Systems Optimization

Alessandra Ghilardi [a,*], Guido Francesco Frate [a], Antonio Piazzi [b], Mauro Tucci [a], Konstantinos Kyprianidis [c], Lorenzo Ferrari [a]

[a] *Department of Energy, Systems, Territory and Construction Engineering, University of Pisa, Pisa, Italy,*

[b] *i-EM s.r.l., Livorno, Italy,*

[c] *Department of Sustainable Energy Systems, School of Business, Society and Engineering, Malardalen University, Västerås, Sweden,*

* alessandra.ghilardi@phd.unipi.it

Abstract

Grid-scale energy storage systems are essential to support renewables integration and ensure grid flexibility simultaneously. As an alternative to electrochemical batteries, Pumped Thermal Energy Storage is a new storage technology suitable for grid-scale applications. This device stores electric energy as thermal exergy, which can be discharged directly for thermal uses or converted back into power depending on the necessities of the grid. The capability of the proposed energy storage to act as electric and thermal storage fits with the sector coupling necessities of multi-energy systems in which electrical and thermal energy carriers are involved. This paper investigates the effects on optimal grid management of integrating a Brayton Pumped Thermal Energy Storage into a multi-energy system. The case study includes renewable generation from photovoltaic modules and residential and industrial users' electrical and thermal load profiles. The system day-ahead optimization, performed through a Mixed Integer Linear Programming approach, aims to minimize the operational cost computed over a 24-hour horizon. The simulation highlights how the proposed storage technology interacts with the users' requirements during different seasons. The final results highlight that using multi-energy storage (i.e., providing power, heating, and cooling) brings a 5% reduction in operating costs during the year compared to a traditional electric-to-electric storage operation.

1. Introduction

Massive exploitation of Renewable Energy Sources (RES) is essential to fulfil the European Union (EU) climate targets for the 2050 net-zero scenario [1]. As a result of the EU policies aimed to face the climate change of the last two decades, many devices for efficiently exploiting RES are nowadays available, such as photovoltaic (PV) modules and wind turbines. Besides this, properly managing and integrating RES into energy systems is essential for reducing carbon emissions.

Strategies for integrating non-dispatchable RES have traditionally focused on the electric grid side since the introduction of the concept of Smart Grids [2]. Despite that, specific operational and planning strategies should address all other energy sectors and their interactions with the electric grid [3]. In this framework, integrated Multi-Energy Systems (MES) can improve the economic and environmental performance of equivalent independent energy systems [4].

MES can include several energy vectors, such as electric, heating, cooling, fuels and transport, who can interact with each other at a district level. MES usually also involve Energy Storage Systems (ESS), essential devices to enhance the system flexibility and fulfil the users' needs [5]. Among the several ESS solutions, multi-energy storages are particularly suitable for MES [6].

Multi-energy storage can store different energy carriers using the same device, thus potentially achieving better economic and environmental performances than separate devices. Carnot Batteries (CBs) are suitable technologies to accomplish this goal since they store electric energy as thermal exergy, which can be directly used or converted back to electricity [7]. CBs are emerging as an alternative grid-scale storage technology due to their long operational life (20-30 years), low cost per kWh [8] and independence from rare raw materials. Since CBs are gaining interest, many technologies have been proposed in the scientific literature for MES optimization, including Liquid Air Energy Storage (LAES) [9],

Compressed Air Energy Storage (CAES) [10], Rankine-based Pumped Thermal Energy Storage (Ra-PTES) [11], and Brayton-based Pumped Thermal Energy Storage (Br-PTES) [12].

Despite these positive features, the economic advantages of using CBs as a pure electric-to-electric storage capacity are still not clearly assessed [13]. However, using CBs in MES as a multi-energy storage capacity could unlock additional revenue streams, improving the CB economic performance.

Among all the CBs technologies, this paper focuses on Br-PTES, given its high round-trip efficiency (50-70%) [14], compared to LAES and Ra-PTES (40-60%) [15] [16]. Br-PTES uses electric energy to power a Br Heat Pump (HP), which charges a High-Temperature Thermal Energy Storage (HT-TES). The stored thermal exergy can then be used directly for heating purposes or as the hot source to power a Br discharging cycle [17]. An additional thermal reservoir, i.e., a Low-Temperature Thermal Energy Storage (LT-TES), can be used to realize a closed-loop configuration [13]. The latter is particularly interesting for MES applications since coupling the electric, heating and cooling networks is a typical requirement at the city-district level [18].

Various storage configurations utilizing solid and liquid media have been proposed [19]. Liquid media, like molten salts (at temperatures between 500-800 K) and cryogenic hydrocarbons (180-300 K), show good resistance to thermal cycles and effective heat transfer capabilities [20]. On the other hand, solid materials such as magnetite, hematite, concrete blocks and ceramic balls tend to be cheaper and can be used in a broader range of operative temperatures [21] when arranged in arrayed packed beds [22].

Although many Br-PTES configurations have been proposed, their integration in MES is barely investigated. Authors in [23] proposed the modelling and integration of Br-PTES at a domestic scale case study which involves different energy vectors. However, Br-PTES achieve higher efficiency when their size is at the grid-scale level, in which they could become cost-effective and competitive with the Li-ion batteries. For these reasons, this paper proposes a novel investigation at a city-district level. Finally, most papers cited in the literature survey have focused on Br-PTES optimal design. However, since the integration of such storage in real systems is recent, the investigation of the optimized management strategies still lacks proper space in the literature.

The contribution of this paper, then, is to simulate a MES which includes electric, heating and cooling loads, RES generation units (PVs), and a Br-PTES to optimize the energy dispatchment at the city-district level. Given this framework, the results highlight how the storage interacts with each energy vector during the different seasons. As a final result, the paper compares the operating costs achieved by the traditional electric-to-electric operation and the multi-energy operation proposed in this study.

2. Methodology

2.1 System architecture

The case study simulates a likely multi-energy system operating at a city-district level located in Sicily, Italy, encompassing different user types, including 35 residential buildings (100 m^2 each) and non-residential buildings, such as 1 supermarket (5000 m^2 of floor area) and 1 healthcare facility (10000 m^2 of floor area). A schematic representation of the district is given in

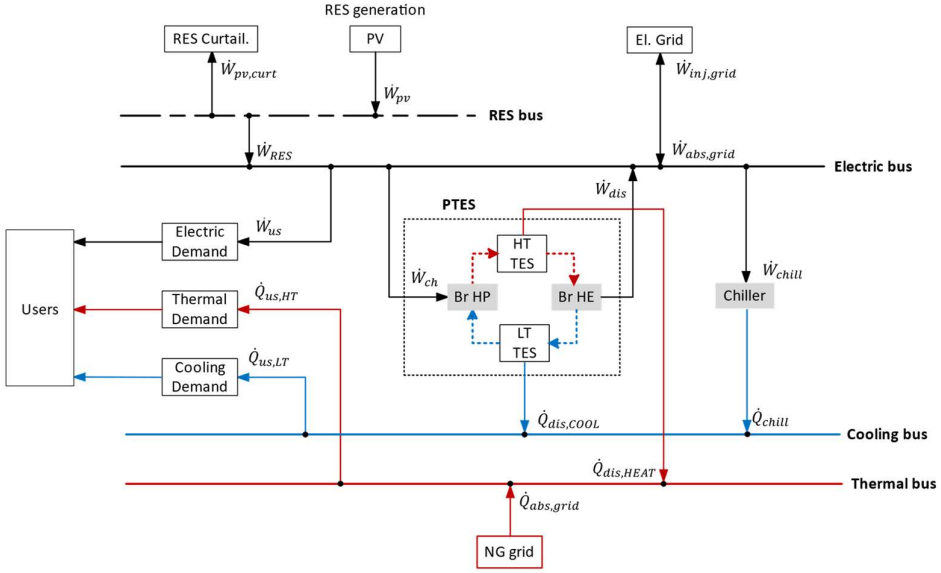


Figure 1. City district model scheme including users' demand, RES production, grid supplies and PTES storage capacity.

Figure 1. Each user is defined by its electric, thermal, and cooling load, with the share of these three vectors varying seasonally. In addition to its reliance on electric and natural gas grids, the system is also equipped with PV modules installed on buildings' rooftops. Finally, the installed storage capacity of the system is provided by a Br-PTES, which can meet the district's electric and heating/cooling necessities.

The electric, heating, and cooling demand profiles are simulated through the software nPro [24]. The software creates the profiles starting from the annual temperature profile (with hourly resolution) of the selected location. By doing so, the electric and thermal profiles are coherent reciprocally and with the outdoor temperature. The latter was provided by the METAR database handled by Iowa Environmental Mesonet [25]. A synthetic overview of the district demands is provided in Table 1.

Concerning the energy prices, ENTSO database [26] and ARERA [27] are used for the cost of absorbed electric energy from the grid, $c_{abs,el}$ measured in €/MWh, and the cost of the absorbed thermal energy from the Natural Gas (NG) grid, $c_{abs,th}$ measured in €/Sm^3.

The PV generation data are simulated starting from the solar radiation data collected from satellite earth observations [28] and processed employing the PVlib library for Python [29] to produce the AC power output. The total installed power of the PV is size 1200 kW, according to the electric and cooling requirements of the district.

Table 1. Electric, heating, and cooling loads of the district

Utility	El. (kW)	Heat (kW)	Cool. (kW)
Residential	112	544	313
Hospital	230	1044	950
Supermarket	69	629	418

2.2 Br-PTES storage

The Br-PTES charging and discharging phases are realized through inverse and direct Brayton-Joule cycles, Brayton Heat Pump (Br-HP) and Brayton Heat Engine (Br-HE), respectively, as represented in Figure 2. Argon is used as the working fluid since it is one of the most common fluids investigated in the literature [30], thanks to its capability of reaching higher temperatures with the same pressure level as other competitive fluids (like helium, nitrogen and air), thus increasing the round-trip efficiency. The HP and the HE operate between a maximum and minimum temperature equal to 500 °C and -80 °C, respectively, thanks to the HT TES and LT TES that act as thermal reservoirs. The storage technology is modelled by defining specific charging and discharging parameters for the HT and LT TES, α_{dis} and α_{ch}, which links the charging and discharging heat flow rates of the TES to the electric charging and discharging net power of charging and discharging phases. Equation (1) and Equation (2) show the definition of these coefficients, where $\bar{\dot{W}}_{ch}$, $\bar{\dot{W}}_{dis}$ are the nominal net power, given by the difference between the compressor charging input power and the turbine discharging output power or vice versa. $\bar{\dot{Q}}_{ch,HT}$ and $\bar{\dot{Q}}_{dis,HT}$ are instead the associated nominal thermal power for charging or discharging the HT TES.

$$\alpha_{ch,HT} = \frac{\bar{\dot{Q}}_{ch,HT}}{\bar{\dot{W}}_{ch}};\ \alpha_{dis,HT} = \frac{\bar{\dot{Q}}_{dis,HT}}{\bar{\dot{W}}_{dis}} \quad (1)$$

$$\alpha_{ch,LT} = \frac{(1-\bar{\dot{Q}}_{ch,HT})}{\bar{\dot{W}}_{ch}};\ \alpha_{dis,LT} = \frac{(1-\bar{\dot{Q}}_{dis,HT})}{\bar{\dot{W}}_{dis}} \quad (2)$$

These coefficients are determined by modelling the charging and discharging thermodynamic cycles by assuming the maximum and minimum cycle temperatures (T_{max} and T_{min}), the iso-entropic efficiency of the compressor, $\eta_{is,c}$, and the turbine $\eta_{is,t}$, and the ratio between the maximum and minimum pressure of the cycles, β, as summarised in Table 2.

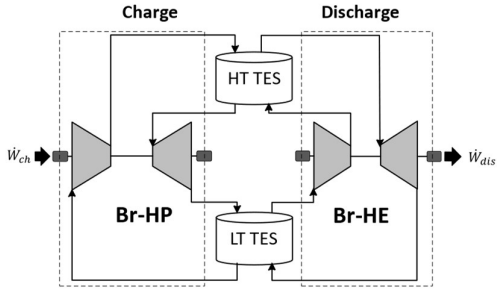


Figure 2. Br-PTES configuration with charge and discharge phases realized by the Br HP and the Br HE, respectively.

The storage model can simulate different operative conditions. Once the storage is charged through electric input by powering the compressor of the HP, the discharge phase can indeed release other vectors, as follows:

- **Electric-to-electric.** The HT and LT tanks act as hot and cold reservoirs to operate a direct Brayton-Joule thermodynamic cycle, which produces electric energy.
- **Electric-to-heating**. The exergy stored in the HT-TES is directly used to fulfil the heating requirements of the district.
- **Electric-to-cooling**. The exergy stored in the LT-TES is directly used to fulfil the cooling requirements of the district.

Table 2. Charging and discharging cycle parameters for the determination of α_{dis} and α_{ch} with Argon as working fluid. Values based on [30].

Parameter	Value
T_{max}	500 °C
T_{min}	-80 °C
β(ch/dis)	4.56/6.54 (-)

$\eta_{is,c}$	0.87 (-)
$\eta_{is,t}$	0.92 (-)

2.3 Sizing of the components

Given the case study's electrical demand and RES production, the storage nominal power ratings, $\bar{\dot{W}}_{ch}$ and $\bar{\dot{W}}_{dis}$, are calculated by a preliminary analysis based on the duration curves of the absolute difference between the electric demand and the RES generation. Particularly, $\bar{\dot{W}}_{ch}$ and $\bar{\dot{W}}_{dis}$ are chosen to match the charging/discharging power required 80% of the time. The charging and discharging durations, τ_{ch}, τ_{dis} are set equal to 6 hours and 3 hours, respectively, which are typical values for RES integration purposes. From these, the nominal HT and LT TES capacities are calculated as in Equation (3):

$$\bar{C}_{HT} = \bar{C}_{LT} = \bar{\dot{W}}_{dis} \cdot \left(\frac{1}{\alpha_{dis}} - 1\right) \cdot \tau_{dis} \quad (3)$$

Concerning the other components of the district, the nominal power absorbed from the electric and NG grids and by the chiller are equal to the maximum of the electric, heating, and cooling demand, respectively, as follows: $\bar{\dot{W}}_{abs,grid} = \max(\dot{W}_{dem})$, $\bar{\dot{Q}}_{abs,grid} = \max(\dot{Q}_{us,HT})$ and $\bar{\dot{Q}}_{chi} = \max(\dot{Q}_{us,LT})$.

Table 3. Storage sizing parameters

Parameter	Value
$\bar{C}_{HT}$	1000 kWh
$\bar{C}_{LT}$	1000 kWh
α_{dis}(HT/LT)	2.7/1.7 (-)
α_{ch}(HT/LT)	1.8/0.8 (-)
$\bar{\dot{W}}_{dis}$	200 kW
$\bar{\dot{W}}_{ch}$	200 kW
τ_{ch}	6 h
τ_{dis}	3 h

2.4 MILP problem formulation

The optimization is realized using a Mixed Integer Linear Programming (MILP) approach, representing state-of-the-art MES optimization techniques. MILP guarantees to find the global optimum in the feasible region Ω, assuming the objective function f_{obj} and the constraints to be linear, and the optimization variables $\boldsymbol{x}$ to be continuous or binary ([0,1] domain). The optimization problem aims to minimize the Operational Cost (OC) and is solved with an hourly timestep t among a 24-hour optimization horizon ($\bar{T}$). The optimization problem is formulated as in Equation. (4), where f_{obj} is given by the sum of the economic losses $\sum_{t=1}^{\bar{T}}\left[\dot{W}_{abs,grid}(t) \cdot c_{abs,el}(t) + \dot{Q}_{abs,grid}(t) \cdot c_{abs,th}(t)\right] \cdot \Delta t$ minus the economic gain $\sum_{t=1}^{\bar{T}}\left[\dot{W}_{inj,grid}(t) \cdot c_{inj,el}(t)\right] \cdot \Delta t$, where $\dot{W}_{abs,grid}$ and $\dot{W}_{inj,grid}$ are the absorbed and injected electric power from the grid, and $\dot{Q}_{abs,grid}$ is the heat flow rate absorbed by the NG grid.

$$\min_{\boldsymbol{x} \in \Omega \in \mathbb{R}^n} \sum\nolimits_{year} (f_{obj}) \quad (4)$$

The optimization algorithm finds the optimal values of the optimization variables $\boldsymbol{x}$, among which the most important ones are:

- $\dot{W}_{ch}$ and $\dot{W}_{dis}$: the charging and discharging power rate for the Br-PTES
- $\dot{W}_{abs,grid}$ and $\dot{Q}_{abs,grid}$: the electrical and thermal heat flow rate provided by the electric and NG grids, respectively.
- The RES curtailment and the PV power input to the electric bus: $\dot{W}_{PV,curt}$ and $\dot{W}_{RES}$, respectively
- Integer variable controlling the on-off status of the storage k_{onoff} (1 is on, 0 is off)
- Integer variable controlling the charging or discharging mode of the PTES, k_{ch} and k_{dis}
- Integer variables controlling the electrical or thermal discharging mode of the storage, k_{el} and k_{th}

Regarding the constraints, the energy balances on the electric, thermal, cooling and RES busses are necessary to guarantee the users' demand fulfilment. Beyond those, some specific constraints characterize the PTES operation. The following list summarizes the most important ones:

- $k_{ch} + k_{dis} \leq k_{onoff,ptes}$. The charging and discharging phases are mutually exclusive (i.e., when the charge is on, the discharge is off, and vice versa)
- $k_{th} + k_{el} \leq k_{dis}$. The electrical and thermal discharges are mutually exclusive (i.e., when the thermal/cooling discharge is on, the electrical discharge is off)
- The State of Charge SOC_k of the k-component (i.e., HT and LT TES) is cyclic over the optimization horizon $\bar{T}$ (Equation 6) and is limited within a $SOC_{k,max}$ and $SOC_{k,min}$ (Equation 5).

$$SOC_{k,min} \leq SOC_k(t) \leq SOC_{k,max} \quad (5)$$

$$SOC_k(t = 0) = SOC_k(\bar{T}) \quad (6)$$

- The charging heat flow rate $\dot{Q}_{ch,k}$ that goes in the HT and LT TES during the charging phase are related to the charging electric heat flow rate, $\dot{W}_{ch}$, by specific coefficients $\alpha_{ch,k}$, which are constant with the HP load, as expressed in Equation (1). The same constraint is valid for the discharging heat flow rates as expressed in Equation (2).

- The SOC of the k-component changes according to the incoming and outcoming heat flow rates, as in Equation (7) and (8), where $\bar{C}_k$ is the TES capacity and $\dot{Q}_{dis,th}$ is the thermal discharge equal to $-\dot{Q}_{dis,HEAT}$ and $+\dot{Q}_{dis,COOL}(t)$ for the HT and LT TES, respectively.

$$SOC_k = SOC_k(t-1) + \Delta SOC_k(t) \quad (7)$$

$$\Delta SOC_k(t) = \frac{\dot{Q}_{ch,k}(t) - \dot{Q}_{dis,k}(t) \pm \dot{Q}_{dis,th}(t)}{\bar{C}_k} \cdot \Delta t \quad (8)$$

- The heat flow rate $\dot{Q}_{chill}$ given by the chiller is related to the electric power absorbed by the electric bus $\dot{W}_{chill}$ by a Energy Efficiency Ratio (EER), which is supposed to be constant with the load. The EER value is set to 3, which is a typical efficiency for chillers for building applications.

3. Results and discussion

3.1 Simulated operation

The optimization process yields the optimal dispatching of the three energy vectors (i.e., electric, thermal, and cooling) the storage delivers. Figure 3 and Figure 4 show a summer and winter representative period, respectively, in which the storage operation faces some typical seasonal patterns of the district energy production and demand. The interaction with the PV production is visible, especially during the summer when the surplus caused by the RES integration is generally used to charge the storage, which is later discharged according to the necessities of the district. For the summer scenario, the electrical discharges usually happen during the first hours of the day, when the air conditioning units of the non-commercial building (supermarket, hospital) are switched on, causing a consumption peak when the PV production is not yet significant. Electrical discharges also happen in the late afternoon when RES production decreases. In that period of the day, the cooling discharges also occur to meet the cooling demand, which is still high (considering the location of the case study). It is worth noting that in the summer period, the storage covers part of the heating load during the central part of the day (required especially in the domestic building for domestic hot water). In this case, indeed, the cooling load is fulfilled by the electric bus, which benefits from the PV production, and the storage then fulfils the heating load to lower its State Of Charge (SOC) and be able to be charged by the surplus in the following timestep. Despite the storage covering part of the cooling load as described, its contribution to the district cooling requirements is limited to a couple of hours in the daytime. This behaviour is because the separate chiller installed in the district works with a higher energy efficiency ratio than the one of the PTES (3 versus 1.7). For this reason, the optimizer chooses to directly exploit the electric grid during the daytime when there is a surplus produced by the RES because it is more convenient and uses the cold stored in the LT TES in the evening hours when the RES production is low.

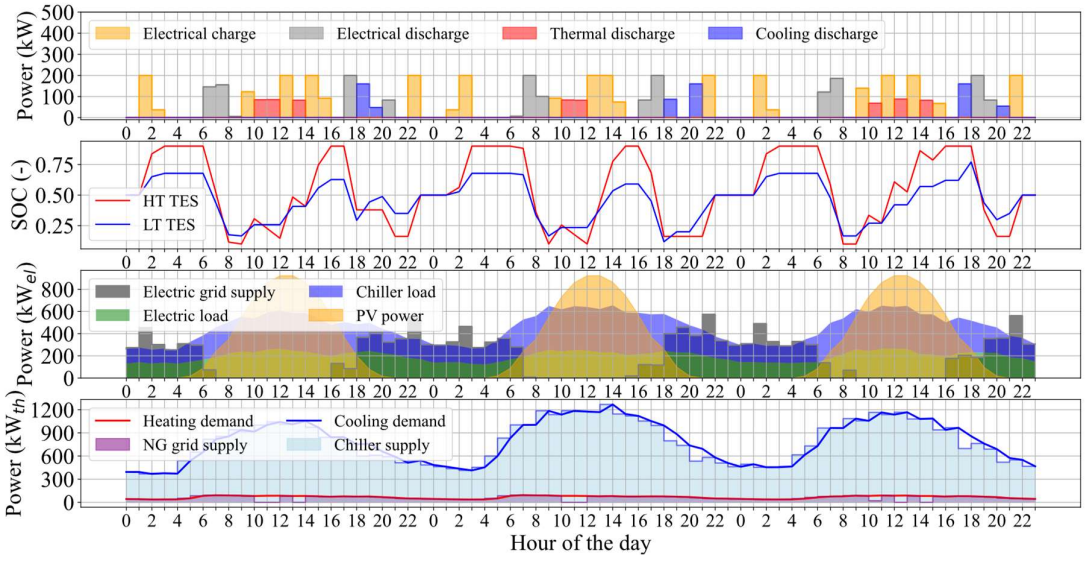


Figure 3. Summer scenario. First image represents the charge/discharge power rates of the storage; second image the SOC of HT and LT storages, third image the electric load fulfillment; forth image the thermal load fulfillment.

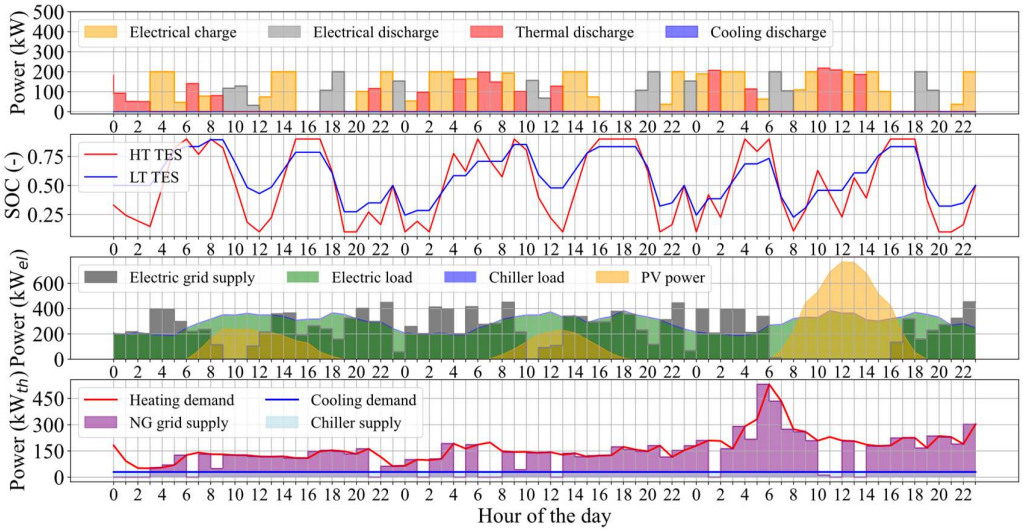


Figure 4. Winter scenario. First image represents the charge/discharge power rates of the storage; second image the SOC of HT and LT storages, third image the electric load fulfillment; forth image the thermal load fulfillment.

A different behavior, instead, happens in the winter scenario. In this case, the PTES significantly contributes to the fulfillment of the heating load, exploiting the HT TES for 5-6 hours per day in the thermal discharge mode. The HT TES, indeed, charges and discharges heat with a higher efficiency compared to LT TES (see values for α_{dis} and α_{ch} in Table 2). Since the LT TES acts as the cold reservoir, indeed, has a limited operation compared to the HT TES, which is the hot reservoir for the involved thermodynamic cycle. This is the reason why the LT TES SOC (visible both in Figure 3 and Figure 4) is not able to follow the HT TES SOC. In other terms, for an equal electrical charge/discharge, the SOC slope of the LT TES is always smaller than the SOC slope of the HT TES. This results in a worse exploitation of the storage capacity, i.e., the SOC of the LT TES varies in a more limited range compared to the HT TES. As a final result, the PTES contribution for the heating is more significant than the one for the cooling.

Besides the RES production, the thermal and electricity prices also affect the storage operation. Both summer and winter scenarios show that sometimes the storage is charged directly from the electric grid during the night-time, buying surplus electricity compared to the electric load. This is due to the lower energy price, which characterizes the night hours. Focusing on the winter scenario, the storage is charged during the night and releases thermal discharge during the first hours of the morning, where the space heating units work with maximum power to heat residential and non-residential buildings. Finally, the electrical discharges happen mostly during the late afternoon, when an electric consumption peak occurs.

3.2 Impact of multi-energy storage operation

The previous section provided a qualitative analysis of the PTES contribution to the analyzed MES, showing a three-day sample period. However, the final purpose of the simulation is to estimate the benefits of using multi-energy storage for a MES application. For this reason, this section provides a quantitative yearly comparison between the PTES operating as a traditional Electric-to-Electric (E-E) storage and the proposed concept of Electric-to-Electric/Heating (E-E/H) or Electric-to-Electric/Heating/cooling (E-E/H/C) storage. Figure 5 compares these three cooperative conditions in terms of OC. The results are normalized with the operating costs in case of no storage capacity installed in the system (OC_0). As the plot clearly shows, the operational costs are reduced when the PTES interacts with the district to provide electricity, heating, and cooling. The only E-E operative condition provides significant cost reduction compared to the no storage case by lowering the OC by 5%. Introducing the E-E/H and E-E/H/C modalities provides an additional

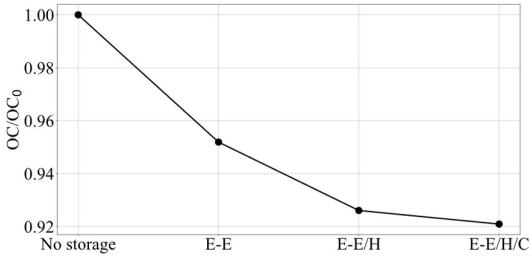


Figure 5. Yearly operating cost for different Operative Costs (OC) of the PTES. Electric-to-Electric (E-E), Electric-to-Electric/Heating (E-E/H) and Electric-to-Electric/Heating/Cooling (E-E/H/C). The subscript 0 refers to the case of no storage capacity.

reduction in the OC, equal to 2% and 1%, respectively. It is worth noting that the E-E/H/C mode brings to a limited improvement of the overall OC, because the PTES works as cold storage for a limited number of hours during the year. This phenomenon occurs due to the limited performance of the LT TES, as discussed in the previous section. Anyway, as a final result, the PTES working as a multi-energy storage is able to reduce the OC of 8% overall.

4. Conclusions

The present work aims to simulate the integration of Brayton-based PTES storage in a multi-energy system at a city-district size, which involves electric, heating, and cooling demand. The system operation was optimized through a Mixed Integer Linear Programming approach to minimize the operating costs. The qualitative results of the simulation showed that the PTES properly interacts with the community by delivering the optimal share of electric, heating, and cooling discharges to fulfil the demand. Beyond that, the quantitative analysis highlighted that using the PTES as a multi-storage capacity significantly reduces the operating costs compared to only electric-to-electric usage, typically proposed for this technology [13]. Comparing these results with the literature, it is found that similar performances (overall operative cost reduction around 10%) were already achieved in [11], by adopting a MILP formulation as well and using a Ra-PTES storage device. However, the cited paper highlights that including the capital costs makes the PTES technology less cost-effective than traditional Li-ion batteries. These findings, then, encourage further analysis to compare the Br-PTES with the other grid-scale technologies, and better define its potential for MES applications. Basing on the benefits showed in the results section, indeed, the proposed storage technology gives support to the idea of enhancing energy communities within multi-energy vector concept, which has been already supported by the new Eu policies.

It is worth to note that the discussed results are related to the single selected case study (i.e., fixed RES, load profiles and storage capacity). Further sensitivity analysis on the storage size, RES penetration and RES profiles (e.g., wind generation beyond solar radiation) could help better define the most suitable application for Br-PTES at a city district level. Finally, from a modelling point of view, this study only considers a first law approach (i.e., only energy exchanges are included in the model). It would then be interesting, as a further step, to evaluate the actual performances of the system when heat is exchanged under temperature differences and assess the potential of Br-PTES with different temperature levels on the demand side.

Acknowledgement

This research has received financial contribution from the Italian Operative National Plan (Piano Operativo Nazionale, PON) in the framework of the project Ricerca e Innovazione 2014−2020 (PON R&I) – Azione IV.6 “Contratti di ricerca su tematiche dell'innovazione e green” (DM MUR 1062/2022).

Project funded under the National Recovery and Resilience Plan (NPRR), Mission 4 Component 2 Investment 1.3 – Call for tender No. 1561 of 11.10.2022 of Ministero dell’Università e della Ricerca (MUR); funded by European Union – NextGenerationEU.

Acronyms and abbreviations

Br	Brayton-Joule
CAES	Compressed Air Energy Storage
CB	Carnot Battery
COOL	Cooling
E-E	Electric to Electric
E-E/H	Electric to Electric/Heating
E-E/H/C	Electric to Electric/Heating/Cooling
EER	Energy Efficiency Ratio
ESS	Energy Storage System
EU	European Union
HE	Heat Engine
HEAT	Heating
HP	Heat Pump
HT	High Temperature
LAES	Liquid Air Energy Storage
LT	Low Temperature
MES	Multi Energy System
MILP	Mixed Integer Linear Programming
NG	Natural Gas
OC	Operating Cost
ORC	Organic Rankine Cycle
PTES	Pumped Thermal Energy Storage
PV	Photovoltaic
Ra	Rankine
RES	Renewable Energy Sources
SOC	State of Charge
TES	Thermal Energy Storage

Symbols

α	Electric to heating/cooling power
β	Compression ratio
c	Cost
$\bar{C}$	Nominal storage capacity
Δ	Difference
f	Function
η	Efficiency
k	Binary optimization variable
Ω	Feasible region
$\dot{Q}$	Thermal power
$\bar{\dot{Q}}$	Nominal thermal power
T	Temperature
$\overline{T}$	Optimization horizon

τ	Duration
t	Time
$\dot{W}$	Electric power
$\bar{\dot{W}}$	Nominal electric power
x	Continous optimization variables

Subscripts

abs	Absorbed
c	Compressor
ch	Charge
chill	Chiller
curt	Curtailment
dem	Demand
dis	Discharge
el	Electric
inj	Injected
is	Iso-entropic
min	Minimum
max	Maximum
obj	Objective
onoff	On-off status
t	Turbine
th	Thermal
us	Users
0	Base-line scenario

References

[1] 2050 long-term strategy n.d. https://climate.ec.europa.eu/eu-action/climate-strategies-targets/2050-long-term-strategy_en (accessed June 5, 2023).

[2] Tuballa ML, Abundo ML. A review of the development of Smart Grid technologies. Renew Sustain Energy Rev 2016;59:710–25. https://doi.org/10.1016/j.rser.2016.01.011.

[3] Mancarella P. Smart Multi-Energy Grids : Concepts , Benefits and Challenges 2012:9–10.

[4] Mancarella P. MES (multi-energy systems): An overview of concepts and evaluation models. Energy 2014;65:1–17. https://doi.org/10.1016/j.energy.2013.10.041.

[5] Victoria M, Zhu K, Brown T, Andresen GB, Greiner M. The role of storage technologies throughout the decarbonization of the sector-coupled European energy system. Energy Convers Manag 2019;201:111977. https://doi.org/10.1016/j.enconman.2019.111977.

[6] Nozari MH, Yaghoubi M, Jafarpur K, Mansoori GA. Development of dynamic energy storage hub concept: A comprehensive literature review of multi storage systems. J Energy Storage 2022;48:103972. https://doi.org/10.1016/j.est.2022.103972.

[7] Dumont O, Frate GF, Pillai A, Lecompte S, De paepe M, Lemort V. Carnot battery technology: A state-of-the-art review. J Energy Storage 2020;32. https://doi.org/10.1016/j.est.2020.101756.

[8] Frate GF, Ferrari L, Desideri U. Energy storage for grid-scale applications: Technology review and economic feasibility analysis. Renew Energy 2021;163:1754–72. https://doi.org/10.1016/j.renene.2020.10.070.

[9] Vecchi A, Li Y, Mancarella P, Sciacovelli A. Multi-energy liquid air energy storage: A novel solution for flexible operation of districts with thermal networks. Energy Convers Manag 2021;238:114161. https://doi.org/10.1016/j.enconman.2021.114161.

[10] Cheng Y, Liu M, Chen H, Yang Z. Optimization of multi-carrier energy system based on new operation mechanism modelling of power-to-gas integrated with CO2-based electrothermal energy storage. Energy 2021;216:119269. https://doi.org/10.1016/j.energy.2020.119269.

[11] Frate GF, Ferrari L, Sdringola P, Desideri U, Sciacovelli A. Thermally integrated pumped thermal energy storage for multi-energy districts: Integrated modelling, assessment and comparison with batteries. J Energy Storage 2023;61:106734. https://doi.org/10.1016/j.est.2023.106734.

[12] Zhang H, Wang L, Lin X, Chen H. Combined cooling, heating, and power generation performance of pumped thermal electricity storage system based on Brayton cycle. Appl Energy 2020;278:115607. https://doi.org/10.1016/j.apenergy.2020.115607.

[13] Mctigue JD, Farres-antunez P, J KS, Markides CN, White AJ. Techno-economic analysis of recuperated Joule-Brayton pumped thermal energy storage. Energy Convers Manag 2022;252:115016. https://doi.org/10.1016/j.enconman.2021.115016.

[14] Steinmann WD, Bauer D, Jockenhöfer H, Johnson M. Pumped thermal energy storage (PTES) as smart sector-coupling technology for heat and electricity. Energy 2019;183:185–90. https://doi.org/10.1016/j.energy.2019.06.058.

[15] Aneke M, Wang M. Energy storage technologies and real life applications – A state of the art review. Appl Energy 2016;179:350–77. https://doi.org/10.1016/j.apenergy.2016.06.097.

[16] Dumont O, Lemort V. Mapping of performance of pumped thermal energy storage (Carnot battery) using waste heat recovery. Energy 2020;211:118963. https://doi.org/10.1016/j.energy.2020.118963.

[17] Olympios A V, McTigue JD, Farres-Antunez P, Tafone A, Romagnoli A, Li Y, et al. Progress and prospects of thermo-mechanical energy storage—a critical review. Prog Energy 2021;3:022001. https://doi.org/10.1088/2516-1083/abdbba.

[18] Ascione F, Canelli M, Francesca R, Masi D, Sasso M, Peter G. Combined cooling , heating and power for small urban districts : An Italian case-study. Appl Therm Eng 2020;71:705–13. https://doi.org/10.1016/j.applthermaleng.2013.10.058.

[19] White A, Parks G, Markides CN. Thermodynamic analysis of pumped thermal electricity storage. Appl Therm Eng 2013;53:291–8. https://doi.org/10.1016/j.applthermaleng.2012.03.030.

[20] Laughlin RB. Pumped thermal grid storage with heat exchange. J Renew Sustain Energy 2017;9. https://doi.org/10.1063/1.4994054.

[21] Zhao Y, Song J, Liu M, Zhao Y, Olympios A V., Sapin P, et al. Thermo-economic assessments of pumped-thermal electricity storage systems employing sensible heat storage materials. Renew Energy 2022;186:431–56. https://doi.org/10.1016/j.renene.2022.01.017.

[22] McTigue JD, Markides CN, White AJ. Performance response of packed-bed thermal storage to cycle duration perturbations. J Energy Storage 2018;19:379–92. https://doi.org/10.1016/j.est.2018.08.016.

[23] Zhang H, Wang L, Lin X, Chen H. Combined cooling, heating, and power generation performance of pumped thermal electricity storage system based on Brayton cycle. Appl Energy 2020;278:115607. https://doi.org/10.1016/j.apenergy.2020.115607.

[24] Wirtz M. nPro: A web-based planning tool for designing district energy systems and thermal networks. Energy 2023;268:126575. https://doi.org/10.1016/j.energy.2022.126575.

[25] Iowa Environmental Mesonet n.d. https://mesonet.agron.iastate.edu/ (accessed June 2, 2023).

[26] ENTSO-E Transparency Platform n.d. https://transparency.entsoe.eu/ (accessed June 2, 2023).

[27] ARERA - Prezzi e tariffe n.d. https://www.arera.it/it/prezzi.htm# (accessed June 2, 2023).

[28] Meteosat series | EUMETSAT n.d. https://www.eumetsat.int/our-satellites/meteosat-series (accessed June 2, 2023).

[29] F. Holmgren W, W. Hansen C, A. Mikofski M. Pvlib Python: a Python Package for Modeling Solar Energy Systems. J Open Source Softw 2018;3:884. https://doi.org/10.21105/joss.00884.

[30] Frate GF, Ferrari L, Desideri U. Techno-Economic Comparison of Brayton Pumped Thermal Electricity Storage (PTES) Systems Based on Solid and Liquid Sensible Heat Storage. Energies 2022;15. https://doi.org/10.3390/en15249595.

Applied Machine Learning for Short-Term Electric Load Forecasting in Cities - A Case Study of Eskilstuna, Sweden

Pontus Netzell[a,*] Hussain Kazmi[b] Konstantinos Kyprianidis[a]

[a]*Future Energy Center, Mälardalen University, Västerås, Sweden* [b]*Dept. of Electrical Engineering, KU Leuven, Leuven, Belgium*
[*]pontus.netzell@gmail.com

Abstract

With the growing demand, electrification, and renewable proliferation, the necessity of being able to forecast future demand in combination with flexible energy usage is tangible. Distribution network operators often have a power capacity limit agreed with the regional grid, and economic penalties await if crossed. This paper investigates how cities could deal with these issues using data-driven approaches. Hierarchical electric load data is analyzed and modeled using Multiple Linear Regression. Key calendar variables holidays, industry vacation, ”Hour of day” and ”Day of week” are identified alongside the meteorological heating-, and cooling degree hours, global irradiance, and wind speed. This inexpensive algorithm outperforms the benchmark ”weekly Naïve” with a relative Root Mean Squared Error of 35% for the year-long rolling origin evaluation. Learnings from the data exploration and modeling are then used to evaluate the AI-based model Light Gradient Boosting Machine. Using similar explanatory variables for this expensive algorithm results in a relative error of 45%, although it outperforms the previous one during the summer. The models have varying strengths and weaknesses and could advantageously be combined into an ensemble model for improving accuracy. Incorporating detailed knowledge of local renewable electricity production in combination with hierarchical forecasting could further increase accuracy. With domain knowledge and statistical analysis, it is possible to create robust load forecasts with acceptable accuracy using easily available machine-learning libraries. Both models have good potential to be used as input to economic optimization and load shifting.

1 Introduction

A part of the solution to reach the global climate goals is to use renewable energy sources, which are volatile, intermittent, and non-dispatchable by nature (Huber et al., 2014). This poses several questions about continued grid stability and conventional power plants need to adapt to this reality by operating more flexibly, ramping up and down at a pace not traditionally seen (Beiron et al., 2020). Uncertainty and volatility in electricity production from variable renewable energy sources could be handled with demand response (Meliani et al., 2021) and the utilization of energy storage for load shifting (Cebulla et al., 2017).

In Sweden, the electrification of the transport and industry sector is crucial for carbon emission reduction, leading to significant growth in electricity demand. Two outstanding examples of industrial growth are the HYBRIT green steel project in the northern parts (Öhman et al., 2022), and the southern Mälardalen region due to its dense population and the addition of new electricity intense industries. Electricity has traditionally been transferred through the national grid from northern hydropower plants, and more recently offshore wind turbines, to the energy-intense southern half of the country. However, due to the rapid growth and end of life for several southern nuclear power plants there are short-term issues in the transfer capabilities, meaning the southern demand cannot be sustainably fulfilled with northern electricity. Intense reinforcement and expansion of the high-voltage grid may eventually make it possible to supply the additional demand. In the long-term however, the increase in electricity usage in the northern areas could lead to a shortage of energy to transfer to the south. Therefore, there is a need for increased local production in the energy-intensive southern cities and regions for a robust and resilient local energy system (Nik et al., 2021).

In Eskilstuna, a city located in the Mälardalen region, the dispatchable local electricity production currently makes up a small portion of the total demand, the rest is imported. The addition of several megawatt-size (MW-size) photovoltaic (PV) parks, and a wind turbine park will increase the yearly energy self-usage ratio. However, it does not resolve the issue on an hourly and seasonal basis, as there is no substantial electricity generation from PV in the evenings as well as during winter in Sweden. With the growing demand, electrification, and renewable proliferation, the necessity of

being able to forecast future demand in combination with flexible energy usage is tangible. Reliable forecasts can enable system operators and utilities to better manage the demand and supply balance in real-time, and control energy storage units for shifting load from high to low production periods, i.e. from day to night, or summer to winter. Use cases for forecasts range from long-term world trends and national changes to medium- and short-term changes on a regional or city-scale level (Hong et al., 2020). Forecasting is essential for the energy and power sector and the area has gotten attention for many decades, but with increasing computational power and new advanced models, the area is regaining focus. Individual investigations are necessary as each dataset is unique and more complex models do not equal increased accuracy. Managing energy assets based on bad forecasts can lead to higher operating costs and, in a worst-case scenario, blackouts in the power grid.

Forecasting can be divided into three main parts using a systems engineering perspective; *Input*, *Model*, and *Output* (Hong & Fan, 2016). Size of historical data for training and the selection of both dependent and independent variables are examples of *Input* variables. If the data is disaggregated by geographical location, then hierarchical forecasting can be chosen as the *Model* technique (Hong et al., 2020). Other *Model* variants are the selection of e.g. non-linear or linear, black-box or non-black-box models, and their respective parameters. The predictions (*Output*) can be combined into ensembles, which is usually considered the best practice (Wang et al., 2018). The application of the forecasts matters, peak prediction generally demands an approach that is different from forecasts used for operational optimization of energy units (Gajowniczek & Ząbkowski, 2017). While numerous forecasting techniques have been proposed, there is no one-size-fits-all, a detailed analysis of the specific case is needed for maximizing the forecast accuracy.

This paper focuses on forecasts and their usage on the urban and sub-urban electricity demand levels in a city via a case study of the Eskilstuna Strängnäs Energi och Miljö (ESEM) electrical grid and energy system. Short-Term Load Forecasting (STLF) is applied to the geographically disaggregated hourly average electric load. The aim of this study is to create and explore a framework to analyze and evaluate forecasting models and determine which calendar and meteorological input variables are best suited for forecasting the electricity demand in cities similar to the studied city. Multiple Linear Regression (MLR) and Light Gradient Boosting Machine (LGBM) are compared to the benchmark ”weekly Naïve” to determine whether advanced AI-based methods provide additional value compared to simpler benchmarks. Implementing these forecasts for control of energy storage units and other flexible assets is discussed, and possible strengths and weaknesses of the two models are emphasized.

The rest of the paper is structured as follows: In section 2: Methodology, data acquisition, algorithm creation, and model selection are presented. In section 3: Results and Discussion, the choice of explanatory variables and model results is presented and criticized. The study is concluded in section 4: Summary and Conclusions, where the road ahead is elaborated.

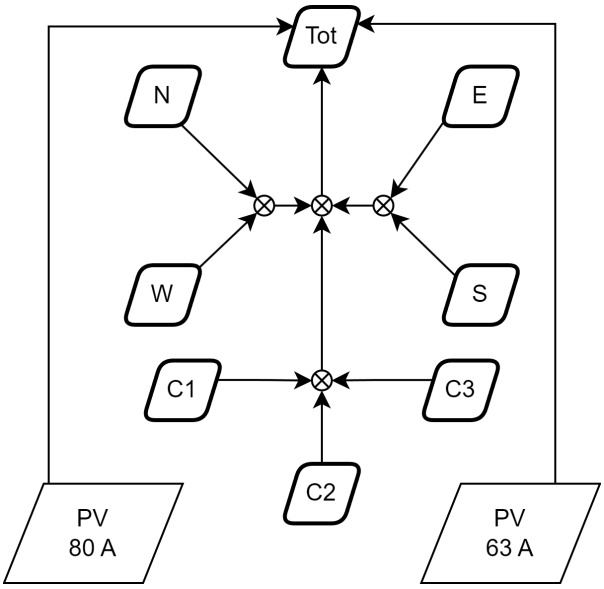


Figure 1. Three central and four outer transformer stations, together with PV-park of different sizes, make up the city’s total energy usage.

2 Methodology

2.1 Data acquisition and pre-processing

The dataset used herein comprises hourly average electrical load in MW from 2020-09-11 to 2022-10-31 (2 years, 1 month, 20 days) and is collected for all entry points (transformer stations) between the regional and the local grid. Local electricity production, e.g. small-scale hydropower generation, has been accounted for according to which transformer station they are connected to. The summation of the seven transformer station loads, together with the generation from all large (63 and 80 Ampere) PV installations, makes up the total energy usage of the city, denoted as ”total energy usage” in this paper (and ”Tot” in Fig. 1). The electricity generation from smaller local PV installations, such as private households, is not included in the total energy usage. The PV installations are not separated into individual time series according to their location; therefore, they must be excluded from the grouped forecasting part of this study. Instead, each individual transformer station and the sum are used for grouped forecasting.

2.2 Data exploration, correlation, and other statistics

To build an accurate forecast model, several meteorological and calendar explanatory variables are evaluated in terms of correlation with total energy usage and improvement in model accuracy. Some of the meteorological variables are reanalysis data of wet and dry temperature, wind speed, rain, and global irradiance from SMHI (2023). Measured in-situ temperature from the central power plant is also used, including smoothed variants, i.e. moving averages with different window sizes. Cross-effects can be calculated by multiplying meteorological and calendar variables (Hong et al., 2010). Degree days and -hours for heating and cooling, which is the temperature difference below or above a certain threshold multiplied by time (Chabouni et al., 2020), are examples of cross effects. The correlation coefficient between each transformer station's load and the reanalysis dry temperature varies between -0.32 to -0.80 (-0.56 for the total load). Such a varying correlation with temperature is indicating the different patterns of usage for different parts of the city. A closer look reveals that the dry temperature gives higher accuracy more often than the wet.

By plotting the load versus different categories, e.g. in a box plot with the hour of the day on the x-axis, the daily load distribution is shown. The load is significantly lower during the night compared to the day. During autumn, winter, and spring a morning peak at 09:00 ±1h, and an afternoon peak at 17:00 ±1h, is identified. However, the load pattern during summer is different, with a single peak at 11:00 ±1h.

Public holidays are considered non-typical days (Eroshenko et al., 2017) where the load is significantly lower. Additive decomposition of the trend, seasonal and residual components (Hyndman & Athanasopoulos, 2018) is applied using the Python library Statsmodels (Seabold & Perktold, 2010). Similar to Işık et al. (2023) the MSTL (Multi Seasonal Trend Decomposition using LOESS (Locally Estimated Scatterplot Smoothing)) reveals daily and weekly seasonality.

2.3 Forecast models and benchmark

The benchmark model is selected as the well-known, in energy forecasting, "weekly Naïve" (copy-paste the previous week's values as the forecast for the next). It captures the weekly seasonality in the data and therefore outperforms the "daily Naïve" (Kolassa et al., 2023). A persistence-based benchmark, meaning finding and copying days that are more similar than simply the weekly pattern, is used in a recent forecasting competition (Farrokhabadi et al., 2022). It can lead to a more accurate Naïve benchmark but at a higher cost of implementation and reduced transferability to other cases, therefore not selected in this study.

The machine learning algorithm MLR is widely used for electric load forecasting and produces forecasts at low computational cost (Kuster et al., 2017). Eq. 1 shows MLR with two independent variables (Hong et al., 2010) as an example:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e \tag{1}$$

where Y is the dependent variable, X_1 and X_2 are independent variables, βs are parameters to estimate, and e is the error term. See Supapo et al. (2017) for a more detailed explanation of MLR. Even though it cannot capture nonlinear relationships by definition, MLR is used because of its scalability and interpretability, while also achieving state-of-the-art performance in many cases. The AI-model LGBM, on the other hand, was highly represented in a recent energy predictor competition (Miller et al., 2020). It is recognized as suitable for electric power modeling, and explained in more detail in the open literature (Tan et al., 2021). One obvious benefit of this model is that it can capture non-linear relationships while still remaining computationally feasible.

The proven track record and community support alongside their simplicity (no hyperparameter tuning), and compatibility (use of the same past and future covariates) conclude that MLR and LGBM are suitable for this comparative study. The models are available in the Python library Darts, which is used in this study (Herzen et al., 2022).

2.4 Algorithm creation

The algorithm (referred to as *Historical Forecasts* in this study) is depicted in Fig. 2. First, necessary inputs are given to the algorithm; forecast horizon, size of historical load for training, number of lagged (past) target values to use, how many hours to jump before making a new prediction, how many predictions to make before retraining, and when to stop. Future and past covariates including their lags can also be given to the model, e.g. temperature and day of the week. A prediction start date is given for splitting the data, otherwise, it will start as soon as possible given the size of the training and available data set. The model is trained, and predictions are made according to the inputs, and at some points retrained. *Historical Forecasts* uses a rolling window approach for the rolling origin evaluation of the forecast (Hewamalage et al., 2023). Each prediction, error, and error metric are saved for further analysis.

2.5 Forecast evaluation: the full-year run

A rolling origin evaluation of the model is applied via the *Historical Forecasts* algorithm using a forecast horizon of 168h, jumping 17h forward between

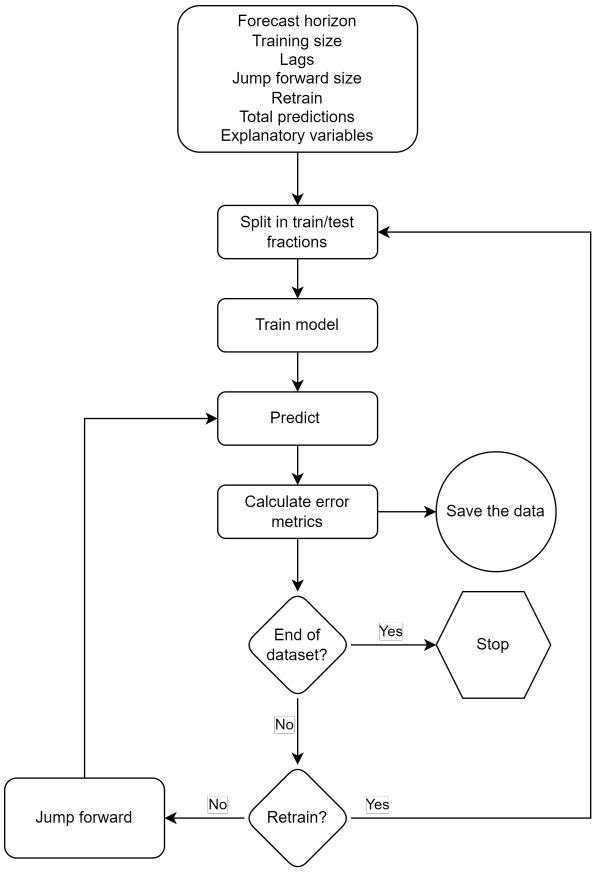


Figure 2. Flowchart of the algorithm *Historical Forecasts* created for this study.

each prediction, retraining every 100th prediction and hence doing a total of 515 predictions. The forecast horizon is selected to match the horizon of available weather forecasts, and the jump between predictions is chosen as a prime number to minimize the chance of resonance with any of the seasonal patterns. Past lags for the load and lags for the future- and past covariates are set to 168h. The evaluated period is approximately 1 year and 1 week, referred to as ”the full-year run”. Given the size of available data, the maximum training size is approximately one year, one month, and two weeks for doing a full-year run. Varying the training size between lower than a year, one year, one year plus two weeks, and maximum size, the ”one year plus two weeks” gave the best accuracy. Including a year of training data and predicting a week ahead means the model has seen the predicted week once, but adding at least one more week to the training data means the predicted week has been seen twice. Including maximum available data showed no significant accuracy improvement. This study is not aiming to prove the optimal training size, as there are many other possible approaches that have not been evaluated. The models are trained four times over the course of the full-year run, as the load profile and temperature dependency is known to be different for the four seasons. During the analysis where the number of retraining was varied, it was shown that re-training too often (every day or week) did not necessarily have a positive effect on accuracy, and certainly not on computational expense. Not re-training at all gave increasingly diverging errors, therefore the final re-training is set to four times. For every full-year run, 86 520 errors (multiply forecast horizon by the total number of predictions) are analyzed, together with 515 average errors (one for each prediction made), and a single average error. Which error metric to be used for different datasets can be derived from Hewamalage et al. (2023). Root-Mean-Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) are concluded as the two most common ones used for STLF of electrical load in Nti et al. (2020), the former used in this study and presented in MW. The relative RMSE (rRMSE), defined as the RMSE of MLR and LGBM respectively divided by RMSE for the ”weekly Naïve”, is used to quantify the performance against the benchmark.

An extensive analysis is done where the least computationally expensive model MLR is used for running hundreds of full-year runs, each generating errors that are compared. Periods with the largest errors, such as public holidays, are focused on separately, as well as the yearly peak, and the summer period. One explanatory variable is added after the other manually, including several combinations, the model parameters are varied, and the results are evaluated. Through combinations of visual inspection of the animations and plotting the model errors in different graphs, calculating and comparing the error metrics, the key explanatory variables are concluded. When no significant improvement is achieved with this semi-structured scrutiny of the MLR, the analysis is stopped. The same analysis is not done with the LGBM due to the computational expense, where only a few selected parameter changes are made to verify the model behavior, e.g. reducing training size reduces accuracy.

3 Results and Discussion

3.1 Final set of explanatory variables

By applying the methodology and analyzing the results, eight explanatory variables are selected, denoted as ”the final set”, shown in Table 1. The impact of adding each explanatory variable to the models is analyzed. The first row of the table shows the average RMSE for the full-year run, including only one explanatory variable; ”Day of Week”. In the second row, the ”Hour of day” is added to the models, and the resulting full-year run RMSE is presented. Consequently, the models in the last row contain all the seven above explanatory variables, including the eighth, Wind speed.

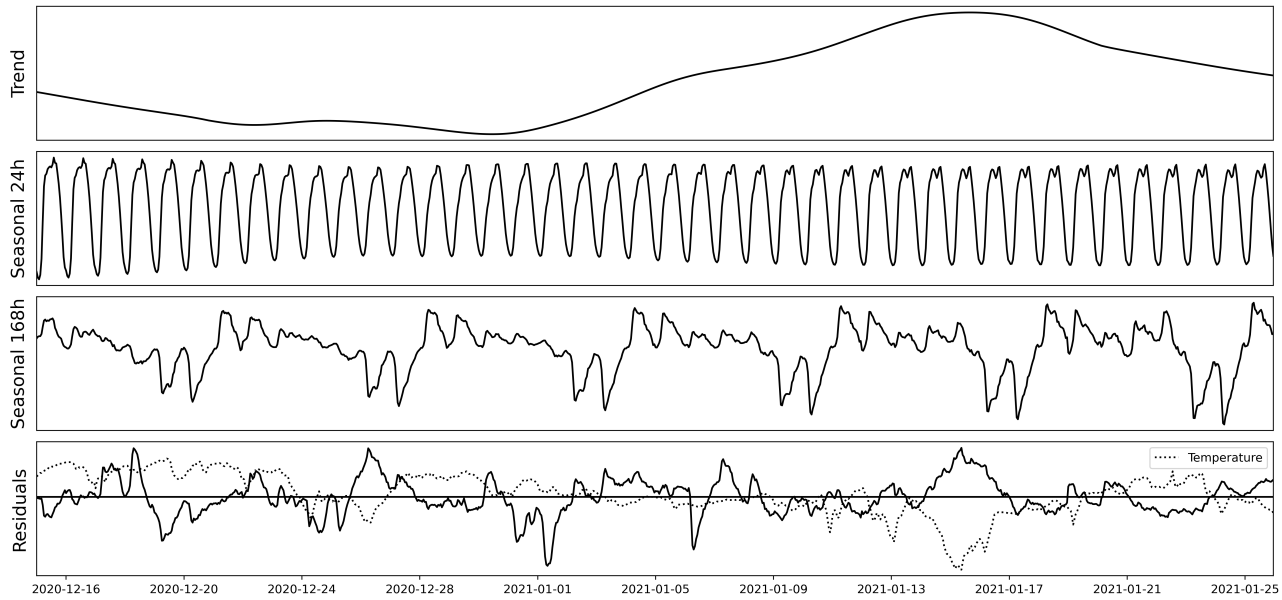


Figure 3. Additive load MSTL with seasonality periods of 24 and 168. Temperature is added to the bottom residuals graph.

Table 1. Final set of explanatory variables and the full-year run average RMSE (MW) for both MLR and LGBM, consecutively adding the explanatory variables in order of appearance

Explanatory variable	MLR	LGBM
Day of week [0-6]	4.54	4.74
Hour of day [0-23]	4.56	4.72
Holidays [0 OR 1]	4.44	4.53
Industry vacation [0 OR 1]	4.31	4.35
Heating hours [Kh, < 10°C]	2.20	2.89
Global irradiance [W/m^2]	2.12	2.68
Cooling hours [Kh, > 20°C]	2.06	2.65
Wind speed [m/s]	2.04	2.71

The choice of the calendar variables ”Day of week” and ”Hour of day” as explanatory variables are justified with the load decomposition, as a daily and weekly seasonal pattern is shown in Fig. 3. Analyzing the bottom residuals graph shows a negative correlation with temperature for this winter example. When not explained by temperature, large peaks in the residuals can be explained with knowledge of public holidays (Christmas and New Year). Further, there is a significant reduction in load due to the common industry practice of closing their operations during the summer vacation period. A binary variable which is set to zero for those four weeks is added, further improving the accuracy shown in Table 1. A variable for covering the thermal load is needed, as electricity is used for heating and cooling. Degree hours are part of the final set, as they give better results than degree days and temperature. Global irradiance and Wind speed improve the accuracy, apart from several of the other meteorological variables, and are therefore included. For justifying degree hours and the use of Holidays, MSTL is applied to the entire dataset, and the residuals are plotted against outdoor temperature in Fig. 4, with public holidays plotted separately. A portion of the residuals are significantly lower than the rest of the residuals during public holidays. Excluding public holidays and adding a LOESS line of the best fit gives a curve explaining how the residuals vary with temperature, depicted as ”Smoothed” in Fig. 4. Residuals are negatively correlated with temperatures below 10°C while positively correlated with temperatures above 20°C.

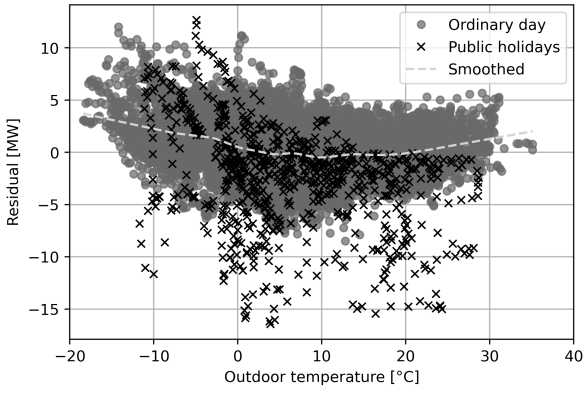


Figure 4. MSTL residuals vs temperature.

3.2 Impact of explanatory variables

Adding certain explanatory variables means only a slight improvement in the full-year run accuracy, and

their existence in the final set needs to be questioned. Adding Wind speed in LGBM reduces large errors for some hours of the year at a cost of a higher average error for the full-year run. Quantifying the economic impact of reducing high errors for a few e.g. windy days, at the cost of a slightly worse overall performance, is a possible way to solidify the existence, and estimate the worth, of the explanatory variables.

The addition of the second explanatory variable ”Hour of day” means that the forecast is performing worse for the MLR as seen in Table 1. However, ”Hour of day” is making the LGBM forecast better and therefore kept in the final set, also because it shows a correlation with the electrical load in the data exploration. Discussions with the stakeholders about the future use of the models can also help in determining whether an explanatory variable should be included in the model.

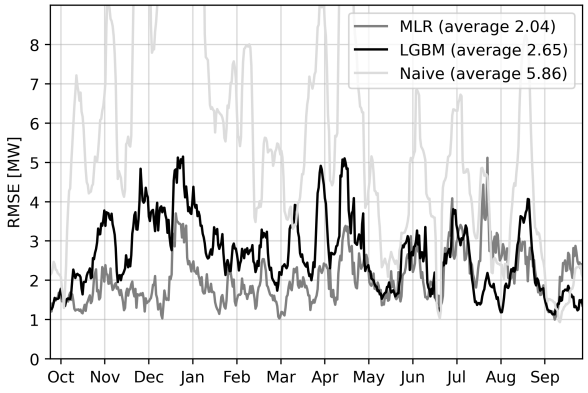


Figure 5. RMSE for MLR, LGBM and ”weekly Naïve”.

3.3 Model results comparison

The results in this paper show that there is no one-size that fits all. When comparing on a single error metric for the full-year runs, the MLR is concluded as superior in terms of accuracy over the LGBM. This despite the fact that a single metric is not giving any detailed insights into the performance of each model. Comparing the performance over the course of the year gives different winners for different periods.

The full-year run results for comparing MLR with the LGBM, including the Naïve benchmark, are shown in Fig. 5. The best-performing models according to Table 1 give an average RMSE of 2.04 (rRMSE of 35%) for MLR and 2.65 (rRMSE of 45%) for LGBM, compared to 5.86 for ”weekly Naïve” in the full-year run. MLR is performing better for three of the four seasons of the year, while LGBM is periodically more accurate during summer, as Fig. 5 shows. A single noon peak during summer, and morning and afternoon dual peaks for the rest of the year in the dataset could be a reason why the LGBM outperforms MLR during periods of the summer and vice versa.

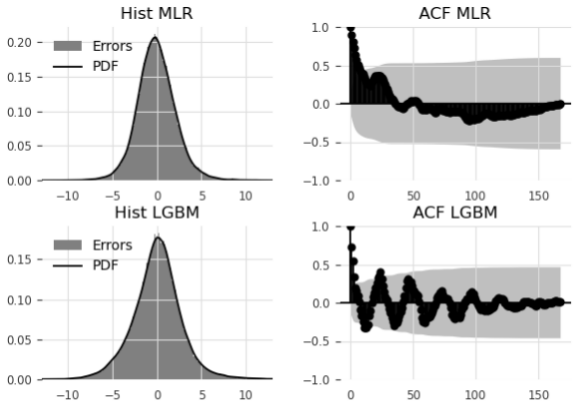


Figure 6. Left) Probability Density Function of the errors. Right) AutoCorrelation Function of the errors.

Another way of comparing the two best-performing models is by analyzing the shape of the histogram of the errors (all 86 520 errors for the full-year run), as seen to the left in Fig. 6. Both models produce errors close to a normal distribution centered close to zero for the full-year run. The centers of the distributions are slightly tilted towards a negative number for MLR, and a positive number for LGBM.

The autocorrelation plots, to the right in Fig. 6, are shown for the same prediction (out of the 515 predictions made, i.e. the 55th) for both models. First, they show that most of the past (lagged) errors are not significantly autocorrelated, except for the first 3–10 errors. This is concluded as serial autocorrelation, meaning if the model is wrong in one direction for the first time step, it will likely be wrong in the same direction in the next step. Second, seasonal autocorrelation is also observed, meaning if the prediction is too low one day, it is likely to be too low on the following day, in a seasonal pattern. Third, these plots highlight that both forecast models produce different errors from an autocorrelation perspective, and are therefore suitable for combining.

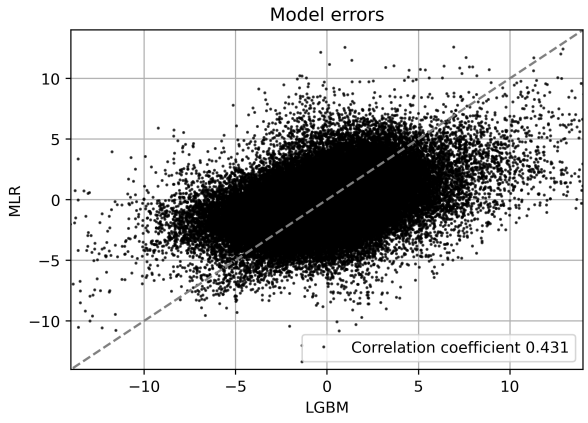


Figure 7. MLR vs LGBM full-year run model errors.

The errors from the two best-performing models are plotted against each other in Fig. 7, and they show a weak correlation. This, together with the distribution of the errors in Fig. 6, shows good potential for combining into an ensemble model. Accuracy improvements are expected when combining models according to the literature (Wang et al., 2018) and forecasting competitions (Miller et al., 2020) but a deeper analysis of this specific case study is needed. The MLR is the winner computationally-wise, it takes about 60 times more time for the LGBM to finish the full-year run.

3.4 Grouped forecasting

The available data are spatially separated and grouped forecasting is applied. The same explanatory variables as the best-performing MLR model and the same settings, e.g. forecast horizon, have been used for the modeling of each individual transformer station. A full-year run is made for all seven transformer stations and the predictions are added together, called Predict Then Sum (PTS). This is compared to the model trained on the sum of the individual transformer stations, called Sum Then Predict (STP), which is slightly different from the total energy usage used in this paper (see Methodology for explanation). The difference between PTS and STP is larger during the heating period, shown in Fig. 8. In general, the grouped PTS forecasting method performs worse, apart from a few exceptions for the full-year run.

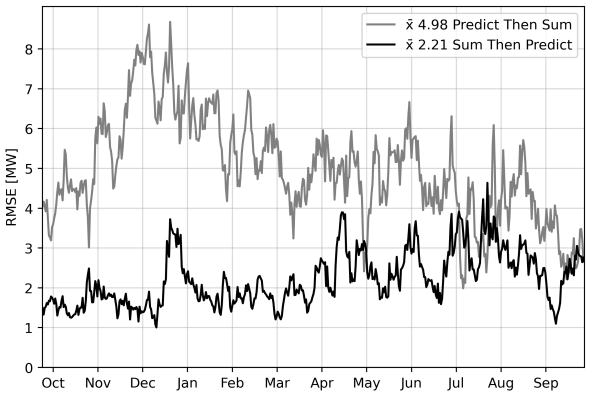


Figure 8. RMSE when forecasting each individual transformer station separately versus their sum.

Comparing PTS with STP show that forecasting on a more aggregated scale is preferred. The correlation between the electrical load and each transformer station, presented in the Methodology section, varies between -0.32 and -0.80. This suggests that the importance of temperature (or temperature variants such as degree hours) when describing the load can vary considerably. Knowledge of each PV park and customized models for each transformer station could improve accuracy further.

3.5 Model dynamics

The algorithm *Historical forecasts* can produce animations of the predictions plotted against the actual load. In Fig. 9, frame 6 out of 25 from an example prediction, which includes the yearly peak of 2021, is shown. The best-performing MLR model is used to produce this frame. The inputs used can be seen at the top left, and the error metrics calculated for this frame are at the bottom left. The bottom error graph shows the difference between the prediction and the actual load for the 168h forecast horizon framed between the two vertical dashed lines. Zooming in around the day of the yearly peak and analyzing the accuracy of the prediction three days before shows that the model underpredicts with approximately 3 MW and 31 MWh for the entire day. If the same analysis is done a few frames later, just 12h before, the numbers change to 1 MW and 10 MWh respectively.

During visual inspection of the predictions, dynamic behaviors are revealed. In some cases, when the model underpredicts the load on the next day, it also underpredicts the day after that, and the following, meaning the average prediction is too low. Dealing with seasonal data makes the presence of seasonal autocorrelation expected. Not until the first day of underprediction has passed does this level out and the model corrects the average level to fit the upcoming days better. This is a good simulation example of how the forecast would have reacted in such a case, it does not know it is underpredicting until the days pass.

Another important performance indicator, which is not straightforwardly easy to measure, is the trustworthiness and explainability of the models. The LGBM produces predictions that do not have a smooth pattern, meaning the first derivative of the predictions during midday alters between positive and negative values consecutively. MLR on the other hand produces predictions where the first derivative less often changes sign and can be seen in Fig. 9. Introducing a forecast model to decision-makers or operators, which don't like or trust it, could affect its usefulness, success, and arguably profitability (Kolassa et al., 2023).

4 Summary and Conclusions

In this paper, a framework to analyze and evaluate forecasting models is explored. The performance of two models, MLR and LGBM, are evaluated using a dataset from the local grid operator of Eskilstuna. Different sets of explanatory- and model variables are tested, concluding the calendar variables; "Day of week", "Hour of day", Holidays and Industry vacation period, and the meteorological variables; Global irradiance and Wind speed together with the cross-effect variables; Heating hours below 10°C and Cooling hours above 20°C as the final set. While

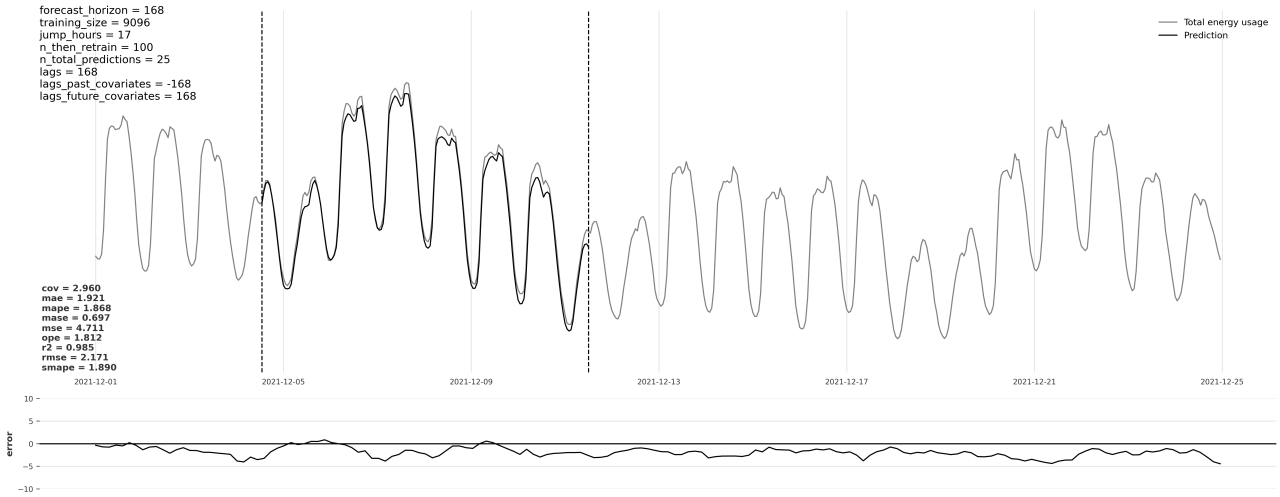


Figure 9. A frame of the animation produced by the algorithm Historical Forecast.

public holidays and non-typical periods still cause the largest errors, adding binary explanatory variables for these significantly improves accuracy. The best-performing MLR and LGBM models outperform the ”weekly Naïve” benchmark model with rRMSE of 35% and 45% respectively. MLR is producing lower errors compared to the more computationally expensive LGBM for the heating period, while it is difficult to unanimously declare a winner for the summer period. Adding an economical dimension can help in determining the acceptable level of accuracy, and suggested measures for enhancing the models is hierarchical and ensemble forecasting. Adding further models (e.g. Artificial Neural Networks based), has the potential to improve the accuracy. Techniques for selecting training data, and optimizing re-training intervals can be investigated further. Expanding the study to include more forecasting models and techniques, and additional explanatory variables (e.g. the national forecast of PV production), would be an interesting path to deepen the knowledge of this specific case.

4.1 The road ahead and future usage

The electrical grid environment is rapidly evolving. Changes in usage patterns, price volatility, and the installation of intermittent renewable energy are just some of the factors that affect the future. In that context, an important aspect when deciding the best model is the ability to adapt and change to better fit the load evolution, adaptability was a key concern during the Covid-19 period (Farrokhabadi et al., 2022). MW-size PV and wind parks are being commissioned from one day to the next, making the historical data less relevant for forecasting. Although important, this study is not evaluating the model’s adaptability, or robustness.
Before implementing forecasts in real-life applications, such as planning and controlling electrical energy storage, an economic dimension should preferably be added to the analysis. Large economic penalties can be the consequence of underpredicting the annual peak on the grid. The best-performing MLR model produced an error in the order of 3 MW for the yearly peak and 31 MWh for that entire day, three days in advance. Depending on the size and availability of local storage, this may or may not be acceptable. This despite the fact that the model accuracy seems to be in line with, or even better than, the overall Swedish national load forecast produced by Svenska kraftnät (Kazmi & Tao, 2022). Forecasting national load is arguably an easier task due to its long history and considerably more in-house knowledge. Internal discussions with the local grid operator of Eskilstuna suggest that the errors are not acceptable, considering possible future energy storage investments. Consequently, point load forecasts, produced here, could be used for everyday short-term management and hourly spot-price optimization, while other methods should be used for peak prediction and peak-shaving as shown in the literature (Gajowniczek & Ząbkowski, 2017).
A future energy storage system will be connected to one of the seven transformer stations. Depending on the location, the forecasting method needs to be considered and customized. This customization, and detailed information about PV parks, can result in high-accuracy forecasts (Hong et al., 2020). Connecting these forecasts to an energy storage management- and sizing problem is the potential next step of this study.

Acknowledgement

The authors hereby thank ESEM for their support, Christer Wiik and Sara Jonsson from the Electrical grid department, and Per Örvind from the Energy department. Thanks to Dr. Moksadur Rahman and Prof. Erik Dahlquist for valuable input and discussions.

References

Beiron, J., Montañés, R. M., Normann, F., & Johnsson, F. (2020). Combined heat and power operational modes for increased product flexibility in a waste incineration plant. *Energy*, *202*, 117696. doi: https://doi.org/10.1016/j.energy.2020.117696

Cebulla, F., Naegler, T., & Pohl, M. (2017). Electrical energy storage in highly renewable european energy systems: Capacity requirements, spatial distribution, and storage dispatch. *Journal of Energy Storage*, *14*, 211–223. doi: https://doi.org/10.1016/j.est.2017.10.004

Chabouni, N., Belarbi, Y., & Benhassine, W. (2020). Electricity load dynamics, temperature and seasonality nexus in algeria. *Energy*, *200*, 117513. doi: https://doi.org/10.1016/j.energy.2020.117513

Eroshenko, S. A., Poroshin, V. I., Senyuk, M. D., & Chunarev, I. V. (2017). Expert models for electric load forecasting of power system. In *2017 ieee conference of russian young researchers in electrical and electronic engineering (eiconrus)* (pp. 1507–1513). St. Petersburg and Moscow, Russia. doi: 10.1109/EIConRus.2017.7910859

Farrokhabadi, M., Browell, J., Wang, Y., Makonin, S., Su, W., & Zareipour, H. (2022). Day-ahead electricity demand forecasting competition: Post-covid paradigm. *IEEE Open Access Journal of Power and Energy*, *9*, 185–191. doi: 10.1109/OAJPE.2022.3161101

Gajowniczek, K., & Ząbkowski, T. (2017). Two-stage electricity demand modeling using machine learning algorithms. *Energies*, *10*(10), 1547. doi: https://doi.org/10.3390/en10101547

Herzen, J., Lässig, F., Piazzetta, S. G., Neuer, T., Tafti, L., Raille, G., … others (2022). Darts: User-friendly modern machine learning for time series. *The Journal of Machine Learning Research*, *23*(1), 5442–5447. doi: 10.48550/arXiv.2110.03224

Hewamalage, H., Ackermann, K., & Bergmeir, C. (2023). Forecast evaluation for data scientists: common pitfalls and best practices. *Data Mining and Knowledge Discovery*, *37*(2), 788–832. doi: https://doi.org/10.1007/s10618-022-00894-5

Hong, T., & Fan, S. (2016). Probabilistic electric load forecasting: A tutorial review. *International Journal of Forecasting*, *32*(3), 914–938. doi: https://doi.org/10.1016/j.ijforecast.2015.11.011

Hong, T., Gui, M., Baran, M. E., & Willis, H. L. (2010). Modeling and forecasting hourly electric load by multiple linear regression with interactions. In *Ieee pes general meeting* (pp. 1–8). Minneapolis, MN, USA. doi: 10.1109/PES.2010.5589959

Hong, T., Pinson, P., Wang, Y., Weron, R., Yang, D., & Zareipour, H. (2020). Energy forecasting: A review and outlook. *IEEE Open Access Journal of Power and Energy*, *7*, 376–388. doi: 10.1109/OAJPE.2020.3029979

Huber, M., Dimkova, D., & Hamacher, T. (2014). Integration of wind and solar power in europe: Assessment of flexibility requirements. *Energy*, *69*, 236–246. doi: https://doi.org/10.1016/j.energy.2014.02.109

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice* (2nd ed.). OTexts: Melbourne, Australia.

Işık, G., Öğüt, H., & Mutlu, M. (2023). Deep learning based electricity demand forecasting to minimize the cost of energy imbalance: A real case application with some fortune 500 companies in türkiye. *Engineering Applications of Artificial Intelligence*, *118*, 105664. doi: 10.1016/j.engappai.2022.105664

Kazmi, H., & Tao, Z. (2022). How good are tso load and renewable generation forecasts: Learning curves, challenges, and the road ahead. *Applied Energy*, *323*, 119565. doi: https://doi.org/10.1016/j.apenergy.2022.119565

Kolassa, S., Rostami-Tabar, B., & Siemsen, E. (2023). *Demand forecasting for executives and professionals* (1st ed.). Chapman and Hall/CRC: Boca Raton, United States.

Kuster, C., Rezgui, Y., & Mourshed, M. (2017). Electrical load forecasting models: A critical systematic review. *Sustainable cities and society*, *35*, 257–270. doi: https://doi.org/10.1016/j.scs.2017.08.009

Meliani, M., Barkany, A. E., Abbassi, I. E., Darcherif, A. M., & Mahmoudi, M. (2021). Energy management in the smart grid: State-of-the-art and future trends. *International Journal of Engineering Business Management*, *13*, 18479790211032920. doi: https://doi.org/10.1177/18479790211032920

Miller, C., Arjunan, P., Kathirgamanathan, A., Fu, C., Roth, J., Park, J. Y., … others (2020). The ashrae great energy predictor iii competition: Overview and results. *Science and Technology for the Built Environment*, *26*(10), 1427–1447. doi: https://doi.org/10.1080/23744731.2020.1795514

Nik, V. M., Perera, A., & Chen, D. (2021). Towards climate resilient urban energy systems: a review. *National Science Review*, *8*(3), nwaa134. doi: https://doi.org/10.1093/nsr/nwaa134

Nti, I. K., Teimeh, M., Nyarko-Boateng, O., & Adekoya, A. F. (2020). Electricity load forecasting: a systematic review. *Journal of Electrical Systems and Information Technology*, *7*(1), 1–19. doi: https://doi.org/10.1186/s43067-020-00021-8

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. In *Proceedings of the 9th python in science conference* (Vol. 57, pp. 10–25080). Austin, TX. doi: 10.25080/Majora-92bf1922-011

SMHI. (2023). *Open data api docs - meteorological forecasts.* Retrieved from `https://opendata.smhi.se/apidocs/metfcst/index.html`

Supapo, K., Santiago, R., & Pacis, M. (2017). Electric load demand forecasting for aborlan-narra-quezon distribution grid in palawan using multiple linear regression. In *2017ieee 9th international conference on humanoid, nanotechnology, information technology, communication and control, environment and management (hnicem)* (pp. 1–6). Manila, Philippines. doi: 10.1109/HNICEM.2017.8269480

Tan, Y., Teng, Z., Zhang, C., Zuo, G., Wang, Z., & Zhao, Z. (2021). Long-term load forecasting based on feature fusion and lightgbm. In *2021 ieee 4th international conference on power and energy applications (icpea)* (pp. 104–109). Busan, Republic of Korea. doi: 10.1109/ICPEA52760.2021.9639313

Wang, Y., Chen, Q., Sun, M., Kang, C., & Xia, Q. (2018). An ensemble forecasting method for the aggregated load with subprofiles. *IEEE Transactions on Smart Grid*, *9*(4), 3906–3908. doi: 10.1109/TSG.2018.2807985

Öhman, A., Karakaya, E., & Urban, F. (2022). Enabling the transition to a fossil-free steel sector: The conditions for technology transfer for hydrogen-based steelmaking in europe. *Energy Research & Social Science*, *84*, 102384. doi: https://doi.org/10.1016/j.erss.2021.102384

Economic investigation of heat pumps for heat recovery from data center

Vahid Zangeneh [a], Lars Erik Øi [a,*]

[a] *Department of Process, Energy and Environmental Technology, University of South-Eastern Norway, N-3901 Porsgrunn, Norway*

lars.oi@usn.no

Abstract

The rapid growth of technology and digitalization lead to an increase in the number of data centers around the world. Data centers produce a considerable amount of heat because of their servers and a large number of electric components. The heat generated by the data centers can be used as a potential source of heating, but the quality (temperature level) of the heat is normally low. In this work, the temperature of the excess (cooling) water from a data center is 45 °C. Generally, there is a possibility to use heat pumps to improve the quality of the heat. To obtain a district heating temperature of 60 °C, 70 °C and 80 °C, the coefficient of performance (COP) was calculated to 5.5, 4.3 and 3.5, respectively. This work is about utilization of the excess heat from a data center with three alternative heat pump solutions with a payback period and economic potential for 10 and 20 years. The simulation process was done by Aspen HYSYS. It was observed that the payback period as expected increases with decreasing COP. The payback period was calculated to values between 2.6 and 5.5 years, depending on the market situation and the delivery temperature. In this work, it is shown that Aspen HYSYS is a reasonable tool to calculate alternatives for heat recovery from data centers based on heat pumps.

Key words: Data center, heat recovery, district heating, heat pump, Aspen HYSYS

1. Introduction

1.1. Background

With the development of information in technology these days, the need for using DCs[1] has steadily increased which has led to an increase in the electricity consumption. On a global scale, the DC electricity demand has risen from about 1.3% of the world's electricity use in 2010 to 2% in 2018 and is expected to keep growing to reach up to 13% in 2030 (Oltmanns *et al.*, 2020). Moreover, the demand for data processing will be increased day to day which means the consumption of higher energy and higher CO_2 emission into the environment and consequently global warming as well as the electricity consumed in a DC almost completely converts to heat. All heat generated by the DCs can be used as a potential source of heating. the cold climate in Nordic countries is extremely suitable for DCs, providing much-needed cooling energy while there is a high demand for heat in these countries. However, the waste heat temperature is generally lower than the required heating temperature, which is a crucial obstacle for using this waste heat.

The heat pump technology can raise the temperature of waste heat to the required heating temperature by using high-grade energy like electricity and make this approach economically, financially, and environmentally profitable. Also, there are many options for waste heat recovery location in DC, such as the waste heat of return air from CRAC[2], CRAH[3], the waste heat of the cooling water from IT room, and the waste heat of cooling water from chillers which will make different capacity and temperature for waste heat recovery. Moreover, required heating temperature and different cycle type of heat pumps can lead to different thermodynamic and financial evaluation (Li *et al.*, 2021). In summary, in this study, thermodynamic and economical evaluation is based on the required heating temperature and a basic cycle of heat pump and then economic consideration is investigated.

1.2. Literature review on energy recovery from DCs

There are different techniques for recycling waste heat for low temperatures. A simple way to reuse

[1] Data centers

[2] Computer room air conditioner

[3] Computer room air handler

low-quality energy is in HVAC[4] or hot water production systems. The temperature of heat waste from air-cooled servers is around 35-45°C. This range is sufficient for reuse heating needs such as domestic heating. By using liquid cooling in DCs it is possible to provide a slightly higher quality of waste heat up to 50-60°C and by two-phase cooling systems, the temperature is as high as 70-80 which is more than sufficient for any heating or hot water application which can be used in DH[5]. Also, this heat provides an income for the DC. District heating is used in Europe more than in the US, particularly in Nordic countries. Moreover, this waste heat can be used for preheating domestic hot water which can lead to energy savings and emission reduction by reducing the use of fossil fuels. And if a higher temperature is required for DH, there is necessary to use heat pumps due to the increase in the temperature of waste heat. The next heat recovery technique is the heating of water in the thermal Rankine cycle of a power plant. The waste heat from the DC is used to preheat boiler feed water which can reduce the consumption of fossil fuels and pollution. They show that it will be more beneficial if a two-phase DC cooling system is used because of the high temperature (Ebrahimi et al., 2014).

Electricity can also be generated by DC waste heat directly through ORC[6] which is investigated by. This technology work as the steam Rankine cycle, but use an organic fluid with a lower boiling point as working fluid. They depict ORC consisting of a turbine, condenser, pump evaporator, and superheater. The superheater is only necessary when the fluid is wet (Chen *et al.*, 2010).

Waste heat of a DC can be used for preheating the water in a coal-fired power plant. One investigation shows that by utilization of waste heat the efficiency of the power plant is increasing up to 2.2% under certain optimized conditions and this performance can lead to a high saving in fuel cost and decrease carbon emission (Marcinichen *et al.*, 2012).

Another investigation is the utilization of excess heat of a DC in the technical university of Darmstadt, Germany (Oltmanns *et al.*, 2020). They show that Direct hot-water cooling for the high-performance computers is provided in the new DC at a temperature 45 °C instead of the current air-cooled servers with water-cooled rear doors at 17-24 °C in the old one. The project shows that between 20-50% of the waste heat generated by high-performance computers can be utilized for heating purposes while the remaining heat is wasted by free cooling. Also, there is a 4% of CO_2 emission reduction on the campus Lichtwiese.

The utilization of DC waste heat for an indoor swimming pool in Barcelona is studied (Oró *et al.,* 2018). Results show that liquid-cooled DC can reduce energy consumption up to 30% in comparison to air-cooled DC.

Waste heat utilization from both the DC and district heating networks in the city Espoo, Finland was investigated. The results showed that the operation cost saving in the system was 0.6 – 7.3%. Also, it was observed that the price of obtained waste heat affects the utilization level of waste heat (Wahlroos *et al.*, 2017).

DC energy efficiency and potential of waste heat capturing analyses showed that waste heat could be captured from 97 % of the total power consumed. Also, it was observed that waste heat from a 1 MW DC could provide the heat demand for over 30,000 m^2 non-domestic building annually (Lu *et al.* 2011).

1.3. Possible temperatures in cooling principle in DCs

Due to proper and efficient utilization of DC waste heat, the temperature of the cooling system not only is very essential but also very sensitive. The quality of heat recovery can be evaluated by the temperature range. Thus, there is some guideline and investigation about the temperature.

One of the important references to determine the favorable environment and temperature and also standard range for DC is provided by the American Society of Heating, Refrigerating, and Air Conditioning Engineers (ASHRAE, 2015). The technical committee recommends that DC equipment should be in the temperature range 18-27°C to fit the manufacturer's provided criteria and also give some information about the allowable range of equipment environmental specifications which shows in Fig. 1. In addition to this, the guideline classify DC to four classes from A1 to A4. Class A1 is a data computer room with tightly controlled environmental parameters such as temperature, dew point, and relative humidity and Class A2/A3/A4 are an information technology space with some controlled environmental parameters.

[4] Heating ventilation air condition.

[5] District heating.

[6] Organic Rankine cycle

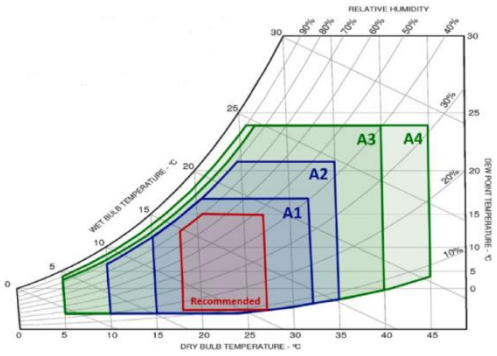


Figure 1: Ambient temperature classification of IT equipment by ASHRAE (Zhang *et al.* 2023)

The temperature of captured waste heat depends on the location where it is captured and on cooling technology. In air-cooled technology, the temperature of captured waste heat is between 25-35 ℃. And by liquid cooling technology waste heat can be captured at a higher temperature between 50-60 ℃ which is better for district heating (Wahlroos *et al.* 2017).

It was observed that the high cooling inlet temperature can be up to 60 ℃ for the water-cooling DC and it means better waste heat utilization (Oltmanns *et al.* 2020).

Based on investigation for having an energy-efficient air cooling (AC) DC, the cold inlet air to all the systems should be maintained at typically 25 ℃ and, output hot air is 40 ℃ and fluidic separation of cold and hot streams is necessary (Patel, 2003).

The optimum temperature range to utilize the waste heat in air cooled DC was shown to be 35-45℃, in water-cooled DC systems and the output temperature could be higher in the range of 60-70 ℃. Also, by two-phase cooling systems, the temperature was as high as 70-80 ℃ (Ebrahimi *et al.* 2014).

Another investigation showed that the inlet temperature of the water could be 60 ℃ to keep junction temperature under 85℃. it showed that around 85% of board heat is collected. Due to providing this criterion, the maximum inlet temperature can be increased to 75 ℃ (Brunschwiler *et al.*, 2009).

2. Methodology

2.1. Process description

Heat pump technology provides an efficient and sustainable solution for utilizing low temperature heating sources. A conventional heat pump is defined as a compression refrigeration cycle powered by either mechanical energy or electricity. In most DC which use waste heat, it is necessary to use a heat pump for increasing output temperature and high quality of waste heat. Ammonia and chlorinated or fluorinated hydrocarbons are usually used in heat pumps as refrigerants. Since chlorofluorohydrocarbons are ozone depleting other refrigerants which are environmentally friendly such as pure hydrocarbons are useful.

The heat pump is made of a number of individual components, including a compressor, a condenser an evaporator, an expansion valve, and a refrigerant circulating from high pressure (red line) to low pressure (blue line). Fig. 2 depicts a mechanical compression of a conventional heat pump. The cooling effect is generated by the cold liquid refrigerant in the evaporator and the heating effect is generated by the hot refrigerant in the condenser.

The refrigerant circulates due to the temperature and pressure difference between the components so that the closed-loop is divided into a high-pressure side and a low-pressure side. A two-phased refrigerant goes into the evaporator where the vaporization of liquid provides the cooling effect and then the refrigerant leaves the evaporator and goes to the compressor. In the compressor, the refrigerant gains high pressure and becomes superheated. The output from the compressor enters to the condenser where the vapoured refrigerant is cooled and condensed to a saturated liquid. In the condenser, the heat of the refrigerant is released to the ambient. After that, the refrigerant enters the expansion valve where it is expanded to lower pressure and the liquid refrigerant is vaporized because of the expansion valve before entering the compressor (Johansson, 2021).

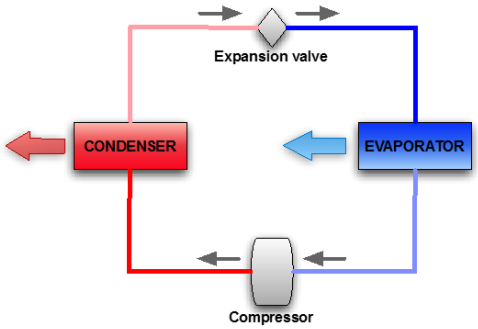


Figure 2: Main components of heat pumps (Øi and Tirados, 2015)

The merit of a refrigerator or heat pump is measured by a parameter called COP. It is the ratio of useful heat given off or taken up by the system to net work done on the system in the one cycle. The equation 1 and 2 represent how to calculate COP of the heat pump (Øi and Tirados, 2015).

$$COP = \frac{Q_{Condenser}}{W} \quad (1)$$

$$W = Q_{condensor} - Q_{Evaporator} \quad (2)$$

In equation 2, $Q_{Condenser}$ is the amount of released heat from the condenser. $Q_{Evaporator}$ is the amount of giving off heat to the evaporator, and W is the power required in the compressor. If there is no heat loss, the difference between input and output heat in the refrigeration cycle is equal to net work of the system.

2.2. Simulation by Aspen HYSYS

Due to calculation and simulation of the cooling system Aspen HYSYS is used. Two pure components, water which is used in the cooling process of DC and refrigerant which is refrig-22 (R-22) are selected in the component list. R-22 is selected as a typical refrigerant, but it is however gradually phased out in the industry due to the ozone depleting effect. After that, Peng-Robinson is selected as a thermodynamic package for simulation in the Aspen HYSYS since it is relevant for these components and applicable for large range of temperature and pressure and two phases, also has a large binary interaction parameter database. The default parameters for the package are used. Then the mechanical equipment of the heat pump which is evaporator, condenser, compressor, and expansion valve is defined with relevant streams.

Three alternative heat pumps are simulated in the Aspen HYSYS. Initial conditions are provided in Tab. 1, 2, and 3.

Table 1: input condition for alternative 1 in Aspen HYSYS

Name	Water 1	Water 6	Fluid 2	Fluid 3
Temperature (^{o}C)	45	60	UN	UN
Pressure (kPa)	101	101	1300	3000

Table 2: input condition for alternative 2 in Aspen HYSYS

Name	Water 1	Water 6	Fluid 2	Fluid 3
Temperature (^{o}C)	45	70	UN	UN
Pressure (kPa)	101	101	1300	3500

Table 3: input condition for alternative 3 in Aspen HYSYS

Name	Water 1	Water 6	Fluid 2	Fluid 3
Temperature (^{o}C)	45	80	UN	UN
Pressure (kPa)	101	101	1300	4000

Modelling of heat pump by Aspen Hysys is presented in Fig. 3.

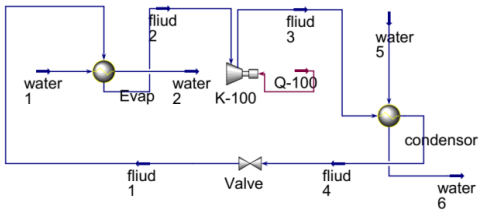


Figure 3: Simulation model of heat pump in Aspen HYSYS

Water 1 is output cooling water from the DC and water 6 is water supplied to the district heating network after using a heat pump.

3. Results

3.1. Simulation results of Aspen HYSYS

The simulation results for three alternatives are shown in Tab. 4, 5 and 6 respectively. As it can be seen, thermodynamic properties of components are calculated by Aspen Hysys software and the heat flow of water and fluid are shown which can be easily calculated the compressor work and condenser heat transfer to calculate COP of heat pumps for three delivery temperatures.

Table 4: 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
V, frac	0	0.32	0	1	1	0	0	0
Temp (°C)	45	33.3	35	33.3	93.1	45	60	69.8
Press (kPa)	101	1300	101	1300	3000	101	101	3000
Mass flow (kg/h)	1000	362.3	1000	362.3	362.3	814.9	814.9	362.3
Heat flow (kJ/h)	-1.580 e7	-2.151 e6	-1.584 e7	-2.108 e6	-2.098 e6	-1.288 e7	-1.282 e7	-2.151 e6

Table 5: Results of material and energy balance achieved from Aspen HYSYS for alternative 2.

	Wat 1	Fluid 1	Wat 2	Fluid 2	Fluid 3	Wat 5	Wat 6	Fluid 4
V. frac	0	0.40	0	1	1	0	0	0
Temp (°C)	45	33.26	35	33.26	104.9	45	70	77.45
Press (kPa)	101	1300	101	1300	3500	101	101	3500
Mass flow (kg/h)	1000	414.5	1000	414.5	414.5	520.4	520.4	414.5
Heat flow (kJ/h)	-1.580 e7	-2.454 e6	-1.584 e7	-2.411 e6	-2.398 e6	-8.223 e6	-8.166 e6	-2.454 e6

Table 6: Results of material and energy balance achieved from Aspen HYSYS for alternative 3.

	Wat 1	Fluid 1	Wat 2	Fluid 2	Fluid 3	Wat 5	Wat 6	Fluid 4
V. frac	0	0.4965	0	1	1	0	0	0
Temp (°C)	45	33.26	35	33.26	115.4	45	80	84.32
Press (kPa)	101	1300	101	1300	4000	101	101	4000
Mass flow (kg/h)	1000	487.6	1000	487.6	487.6	400.4	400.4	487.6
Heat flow (kJ/h)	-1.58 e7	-2.88 e6	-1.584 e7	-2.836 e6	-2.819 e6	-6.327 e6	-6.266 e6	-2.88 e6

3.2. Calculation of COP for heat pump

After simulation by Aspen Hysys, the COP of three alternative Heat Pumps is calculated.
For alternative 1:

$$COP = \frac{Q_{condenser}}{W} = \frac{52800}{9687} = 5.45$$

For alternative 2:

$$COP = \frac{Q_{condenser}}{W} = \frac{56310}{13150} = 4.282$$

For alternative 3:

$$COP = \frac{Q_{condenser}}{W} = \frac{60740}{17580} = 3.455$$

So, when the supplies water to district heating are 60, 70 and 80 the COP are 5.45, 4.282 and 3.455 respectively.

3.3. Economic calculation

To calculate the energy cost, it is used simple assumptions. Economic calculation is done for two market situations. In the first market situation, electricity price is assumed 0.107 EUR/kWh in winter and 0.05 EUR/kWh in summer and in the second one the electricity price is equivalent 0.107 EUR/kWh or 1.07 NOK/kWh during the year, and the district heating price is obtained from DH company which is 0.05 EUR/kWh (Li *et al.*, 2021). Calculation of economic potential is presented by equation 3.

$$Economic\ potential = Price \times Recovered\ heat - \left(\frac{El\ Price \times Recovered\ Heat}{COP}\right) \quad (3)$$

One 7 MW DC is investigated, and 100% heat recovery is assumed, so the recovered heat is calculated to 60 GWh/yr in the DC. The economic potential for no heat pump and three alternative heat pumps by the first economic market situation are calculated below.

$$Economic\ potential\ without\ heat\ pump = Price \times Recovered\ heat = 0.05\frac{\text{EUR}}{\text{kWh}} \times 60\ \frac{\text{GWh}}{\text{yr}} = 3.00\ \frac{\text{MEUR}}{\text{yr}}$$

$$Economic\ potential\ with\ heat\ pump\ 1 = 0.05\frac{\text{EUR}}{\text{kWh}} \times 60\ \frac{\text{GWh}}{\text{yr}} - \left(\frac{(\frac{0.107+0.05}{2})\frac{\text{EUR}}{\text{kWh}} \times 60\ \frac{\text{GWh}}{\text{yr}}}{5.45}\right) = 2.136\ \frac{\text{MEUR}}{\text{yr}}$$

$$Economic\ potential\ with\ heat\ pump\ 2 = 0.05\frac{\text{EUR}}{\text{kWh}} \times 60\ \frac{\text{GWh}}{\text{yr}} - \left(\frac{(\frac{0.107+0.05}{2})\frac{\text{EUR}}{\text{kWh}} \times 60\ \frac{\text{GWh}}{\text{yr}}}{4.282}\right) = 1.9\ \frac{\text{MEUR}}{\text{yr}}$$

$$Economic\ potential\ with\ heat\ pump\ 3 = 0.05\frac{\text{EUR}}{\text{kWh}} \times 60\ \frac{\text{GWh}}{\text{yr}} - \left(\frac{(\frac{0.107+0.05}{2})\frac{\text{EUR}}{\text{kWh}} \times 60\ \frac{\text{GWh}}{\text{yr}}}{3.455}\right) = 1.637\ \frac{\text{MEUR}}{\text{yr}}$$

For the case of no heat pump, all energy of the DC is utilized which is worth 3 MEUR while for three alternative heat pumps the economic potential decreased.
Economic potential for heat pumps also are calculated by the second market situation and the results are 1.82 MEUR/yr, 1.5 MEUR/yr and 1.14 MEUR/yr, respectively. However, If there is cheap surplus renewable energy available in special circumstances, it may improve the economy compared to the calculation.

3.4. Investment cost

The total investment cost for heat pumps based on excess heat as heat source according to all categories are shown Tab. 7 (Li *et al.*, 2021).

Table 7: total investment cost for HP project (Li *et al.*, 2021)

HP capacity	Specific cost, million €/MW
0.5 MW< HP capacity<1 MW	1.3 to 0.97
1 MW< HP capacity<4 MW	0.97 to 0.72
4 MW< HP capacity<10 MW	0.72 to 0.67

Therefore, the investment cost of a heat pump in the 7 MW DC is in the third category, it is between the amount of 0.72 to 0.67 M€/MW, and for simplicity 0.7 M€/Mw is assumed in the investigation. Since the DC is assumed 7 MW the total investment cost is 4.9 M€.

3.5. Calculation of the payback period

The payback period is the time that the initial investment is fully recovered. The payback period PB is calculated in equation 4 (Li *et al.*, 2021).

$$B_{sav}\left(\frac{(1+i)^{PB}-1}{i(1+i)^{PB}}\right) - Invt = B_{sav}\left(\frac{1-(1+i)^{-PB}}{i}\right) - Invt = 0 \quad (4)$$

In the equation, B_{sav} is the annual energy bill saving, Invt is the initial investment. The interest rate is i which in this study is 7 %. The payback period, PB, indicates the number of years for the recovery of the investment.

The payback period is calculated for three delivery temperatures by two market situations and the results are shown in Fig. 4 so that when the delivery temperature goes down or the COP of heat pumps goes up, the payback period reduces.

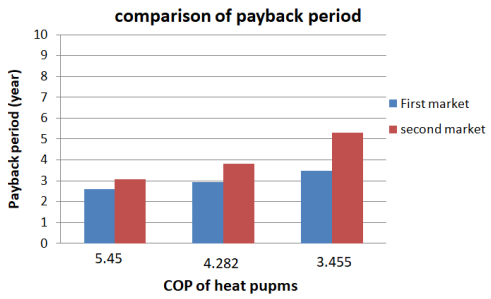


Figure 4: Comparison of payback period for three delivery temperatures by two market situations

3.6. Economic result

The economic result for all four alternatives is calculated when the project are run for 10 and 20 years. The factor for constant income is given by equation 5.

$$Factor = \left(\frac{1-(1+i)^{-n}}{i}\right) \quad (5)$$

For n=10 years and i=7% the factor is 7.02, and for n=20 years and i=7% the factor is 10.59. Therefore, the economic result for the four alternatives is calculated by equation 6 (Sharfuddin and Øi, 2020).

$$Economic\ result = (Economic\ potential \times factor) - Investment\ cost \quad (6)$$

It is noticeable that for the case without heat pump, there is no investment cost for installing heat pump. Therefore, the economic result for 10 years is 21.06 MEUR and for 20 years is 31.77 MEUR. Also, economic result is calculated for three delivery temperatures based equation 6. The results of 20 years investigation are shown in Fig. 5, 6 for all alternatives and by the first and second market situations respectively, so that there is an increased trend of economic potential by increasing COP of heat pump or decreasing delivery temperature. Also, the same pattern is for 10 years calculation.

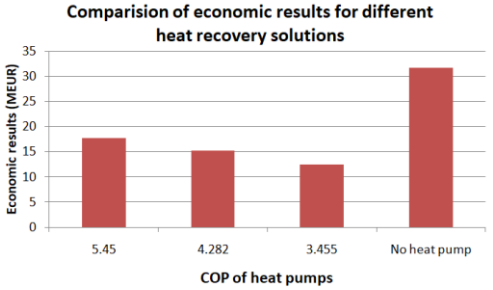


Figure 5: Comparison of economic result of investigation in 20 years by first market situation

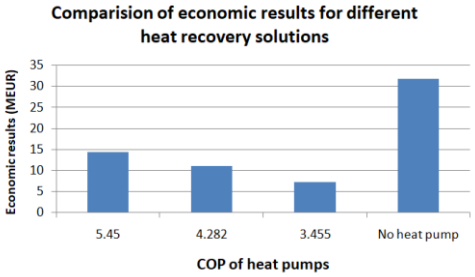


Figure 6: Comparison of economic result of investigation in 20 years by second market situation

4. Summary and Discussion

The excess heat of DCs can be utilized as a renewable source of energy. Due to having efficient utilization of excess heat DC, the excess heat should be connected to a district heating network. However, the quality of heat from the DC is low and needs to be improved. A heat pump can be used for the purpose of improving heat quality. Aspen HYSYS software is used for simulation and economic optimization under different conditions. In all cases, water will leave the DC at 45℃, and in one case without a heat pump and three alternatives with heat pumps entering the district heating network at 60℃, 70℃ and 80℃ are simulated. After simulation by Aspen Hysys, it is observed that the COP of three alternative heat pumps decreases by increasing water supply temperature (5.5, 4.3, and 3.5 respectively).

Economic considerations including investment cost, payback period, and the economic result for 10 and 20 years have been done by two market situations with specified prices for electricity and district heating costs.

It is observed that the payback period as expected increases by decreasing COP and higher supply water temperature so that it changes from 2.6, 2.9, and 3.5 years to 3.1, 3.8, and 5.3 years for the three delivery temperatures with the first and second market situations of calculation respectively.

Moreover, the economic potential for three delivery temperatures show that a higher COP will produce a higher economic potential and a lower COP will give lower economic values. Also, it is shown that the first market situation provides higher economic potential when the price of electricity cost has been calculated in two parts.

However, it is noticeable that the same price for district heating cost is assumed for all alternatives while the water supply temperature using a heat pump is higher than without a heat pump. That is why without heat pump scenario is the most economical alternative for two market situations. Also, the price of electricity and district heating may vary from one place to another which can affect the results. Another item that can affect the results is pipeline cost for the district heating network which has a dependency on climate, length of connection, and environmental situation. All in all, the calculation and investigation depict that a high potential for utilization of waste heat is available from DC.

References

ASHRAE handbook. (2015) ‘Thermal guidelines for data processing environments–expanded data center classes and usage guidance’, *Whitepaper prepared by ASHRAE technical committee (TC), Atlanta, USA*

Brunschwiler, T. *et al.* (2009) ‘Toward zero-emission data centers through direct reuse of thermal energy’, *IBM Journal of Research and Development*, 53(3), pp. 1-13, doi: 10.1147/JRD.2009.5429024

Chen, H. *et al.* (2010) ‘A review of thermodynamic cycles and working fluids for the conversion of low-grade heat’, *International Journal of Renewable and Sustainable Energy Reviews*, 14, pp. 3059-3067. doi: 10.1016/j.rser.2010.07.006

Ebrahimi, K. *et al.* (2014) ‘A review of data center cooling technology, operating conditions and the corresponding low-grade waste heat recovery opportunities’, *International Journal of Renewable and Sustainable Energy Reviews*, 31, pp. 622-638. doi: 10.1016/j.rser.2013.12.007.

Johansson, P. (2021) ‘Heat pumps in Sweden - A historical review’, *International Journal of Energy*, 229(120683), doi: 10.1016/j.energy.2021.120683.

Li, H. *et al.* (2021) ‘Energy, economic, and environmental analysis of integration of thermal energy storage into district heating systems using waste heat from data centres’, *International Journal of Energy*, 219(119882) doi: 10.1016/j.energy.2020.119582.

Li, J. *et al.* (2021) ‘Optimal schemes and benefits of recovering waste heat from data center for district heating by CO2 transcritical heat pumps’, *International Journal of Energy Conversion and Management*, 245 (114591), doi: 10.1016/j.enconman.2021.114591.

Lu, T. *et al.* (2011) ‘Investigation of air management and energy performance in a data center in Finland: case study’, *International Journal of Energy and Buildings*, 43, pp. 3360-3372, doi: 10.1016/j.enbuild.2011.08.034.

Marcinichen, J.B. *et al.* (2012) ‘Potential for waste heat utilization of hot-water-cooled data centers: A case study’, *International Journal of Applied Thermal Engineering*, 41, pp. 36-51. doi: 10.1016/j.applthermaleng.2011.12.008.

Øi, L. E. and Tirados, I. Y. (2015) ‘Heat pump efficiencies simulated in Aspen HYSYS and Aspen Plus’, *Proceedings of the 56th Conference on Simulation and Modelling, SIMS 2015*, Scandinavian Simulation Society, doi: 10.3384/ecp15119141.

Oltmanns, J. *et al.* (2020) ‘Potential for waste heat utilization of hot-water-cooled data centers: A case study’, *International Journal of Energy Science and Engineering*, 7, pp.1793-1810. doi: 10.1002/ese3.633.

Oró, E. *et al.* (2018) ‘Design and economic analysis of liquid cooled data centres for waste heat recovery: A case study for an indoor swimming pool’, *International Journal of Sustainable cities and society*, 36, pp. 185-203. doi: j.scs.2017.10.012.

Patel, C. D. (2003) ‘A vision of energy aware computing from chips to data centers’, *The international symposium on micro-mechanical engineering, ISMME2003.*

Sharfuddin, M. and Øi, L. E. (2020) ‘Simulation of heat recovery from data centers using heat pumps’, *61th International Conference of Scandinavian Simulation Society, SIMS 2020*, Virtual Conference, Finland doi: 10.3384/ecp2017671.

Wahlroos, M. *et al.* (2017) ‘Utilizing data center waste heat in district heating – impacts on energy efficiency and prospects for low-temperature district heating networks’, *International Journal of energy*, 140, pp. 1228-1238, doi: 10.1016/j.energy.2017.08.078

Zhang, Y. *et al.* (2023) ‘Research and Technologies for next-generation high-temperature data centers – State-of-the-arts and future perspectives’, *International Journal of Renewable and Sustainable Energy Reviews*, 171 (112991), doi: 10.1016/j.rser.2022.112991.

Design of Machine Learning method for decision-making support and reliability improvement in the investment casting process

Antonia Antoniadou [a,b], Konstantinos Kyprianidis [a], Ioanna Aslanidou [a], Anestis Kalfas [c] and Dimitrios Siafakas [b]

[a] *Mälardalen University, Vasteras, 72220, Sweden,* [b] *TPC Components AB, Hallstahammar,734 92, Sweden,* [c] *Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece*

antonia.antoniadou@mdu.se

Abstract

The need to improve reliability and support decision-making in manufacturing has drawn attention to the application of diagnostic and decision-support tools. Particularly in the investment casting industry, data-driven methods can be the enabler for process diagnostics and decision support. Images from the microscopic examination in the investment casting process are used as data input, to detect defects in produced pieces. The microscopic examination usually relies solely upon the ability of the operator to determine whether an image from the microscope contains a defect. Therefore, an effective strategy for this decision-making process is crucial to improve the reliability of the examination. The use of the machine learning classifier Random Forest is introduced to derive predictions on the existence of a defect in the input image. This work focuses on employing machine learning tools for image recognition and the developed approach constitutes a decision support model to assist the operator and improve the reliability of their assessment.

1. Introduction

During the last decade, machine learning (ML) techniques have been widely implemented in different production processes, aiming to enhance the quality of the products, apply process diagnostics, or support decision-making (Esmaeilian et al., 2016) Utilization of ML methods has found application in production operational management centers to facilitate decision-making processes (González Rodríguez et al., 2020), or use predictions to support decisions in inventory management (Mohamed & Saber, 2023). The need to improve the reliability of decision-making for fault detection and diagnostic processes represents one of the strategic objectives of many industries. In manufacturing, reliability refers to machines, equipment, and systems being able to perform their intended functions with consistency and predictability. Providing reliable products is vital to the success of the industry, as traditionally reliability is evaluated by the final product quality (Safhi et al., 2019). Numerous measures can be taken to increase manufacturing reliability, such as regular maintenance and calibration of equipment, as well as diagnosing faults in components or systems.

The microscopic examination mentioned in this work is a part of the investment casting process, a process aiming to create components that can be used in turbomachinery applications, characterized by high geometrical complexity, and later subjected to demanding performance conditions. The production of such parts has multiple subprocesses and is a very sophisticated procedure with much attention to detail (Warren et al., 2021).

Most current practices in industry involve experts inspecting individually each piece produced and detecting defects manually (Jawahar et al., 2021). Particularly in the aerospace manufacturing industry, visual inspection still dominates the testing of parts including engine blades, accounting for approximately 90% of all inspections (Aust et al., 2021). With quality assessment being one of the essential steps of the process, relying solely on the ability of an inspector to detect faults could be of high risk (Aust et al., 2021). Studies have shown that during the inspection of parts, the judgment of professionals can be biased by expectations coming from contexts such as prior knowledge or experience and inspectors may be unaware when their judgments are affected (MacLean & Dror, 2021). Bias can come from different sources, either case-specific, such as data, reference materials, and contextual information, depending on the environment and experience, or cognitive architecture and human nature. Many studies have so far been carried out on using ML techniques to identify faults and improve the reliability of other processes such as fluorescent penetrant inspection

(Niccolai et al., 2021) (Shipway et al., 2019), or X-ray inspection (Jiang et al., 2021), (García Pérez et al., 2022), but little has been done on the microscopic examination of the parts, and even less on the investment casting products.

The microscopic examination process in investment casting appears to be an excellent opportunity for the application of ML methods that could improve the reliability of fault diagnosis by assisting in decision-making. This is due to the requirement that the inspector conducting the examination detects discontinuities in materials and decides whether they could endanger the structural integrity of the produced part and its functionality. The purpose of this research work is focused on improving the reliability of the decision-making mechanism of the inspector's assessment during the microscopic examination, through the application of ML techniques. As it is essential to reduce the risk of false assessment when diagnosing faults while minimizing possible bias and increasing objectivity, an assisting ML model for the operator is proposed.

2. Methodology

2.1. Background

Investment casting is a manufacturing process that produces dimensionally accurate components and is a more cost-effective alternative to forging or machining since waste materials are reduced to a very low level (Li & Wang, 2021). During the process, molten wax is injected into a metallic mold to create a wax pattern with the desired component shape. The wax mold is repeatedly dipped into a ceramic slurry which then hardens to create a ceramic casing around the wax design. The wax is then removed from the shell by melting, leaving a cavity inside that exactly resembles the shape of the component. The casting procedure itself is carried out by filling the thus-produced ceramic shell with molten alloy after hardening the ceramic shell by heating. The shell is separated as the molten metal hardens to produce the components which will then undergo various finishing and inspection processes (Del Vecchio et al., 2019).

As one of the final inspection methods, the microscopic examination contains assessments whose results determine if a part is ready for delivery, based on the requirements of each customer. Usual requirements might be the maximum allowed percentages of a specific defect found on a part, such as porosity. The conditions of the casting process in its entirety strongly determine the occurrence of defects during the observation. The operation of microscopic examination within the factory typically relies solely upon the ability of the operator, without any assisting model. The inspection is carried out using portable equipment and conventional optical microscopy procedures. For the microscopic examination to be successful and with accurate results, the operator is required to search for and detect irregularities by visually examining the cut surface of the material, an example of which is shown in Figure 1.

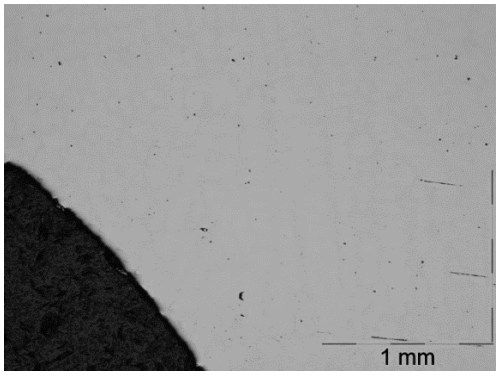


Figure 1: The cut up of a piece produced by investment casting.

When inspecting the material in the microscope, the operator comes across images that are either clean or contain a defect. It is up to the operator to decide the status of every image (faulty or non-faulty). Faults appearing on the image can be:

- Porosity (often forming as microporosity), which appears in the microscope as dark repeated streaks with smooth edges. It is known to be the most common defect found during investment casting and dramatically limits the life of aerospace components (Torroba et al., 2014). A possible porosity cause is shrinkage during solidification, where there is a shortage of molten material flow in-between the space of connected dendrites (Mozammil et al., 2020).
- Gas pockets, that are observed as dark circular shapes during the microscope examination. Similar to porosity, factors such as metal solidification time and air entrapment are due to turbulence during the pouring of the molten metal into the shell (Kaiser et al., 2011).
- Cracks, which are usually caused by internal stresses from the solidification of the metal or rapid cooling, can be identified as either hot tear cracks, appearing as noncontinuous dark

lines of variable widths, or cold state, which indicates thin continuous lines. They can also initiate forming by other defects or intermetallics (Dezecot & Brochu, 2015).

- Chemical reactions that might appear due to refractories used during the forming of the shell mold (Hao et al., 2020). They can be a result of the interaction of the metal used in investment casting and the ceramic mold where the metal is poured.

Examples of the above are shown in Figure 4. There are also fewer common cases of other defects appearing such as missruns, dross, or segregation, generated mainly during the solidification process. During the microscopic assessment, the inspector may come across one of the defects or possible combinations of them.

2.1. Approach

The method proposed in this paper aims to detect faults that can occur during the investment casting process and support the assessment of defect presence. The steps followed in this work to tackle the risk of inaccurate assessment and create an assisting model for the operator were designed as follows:

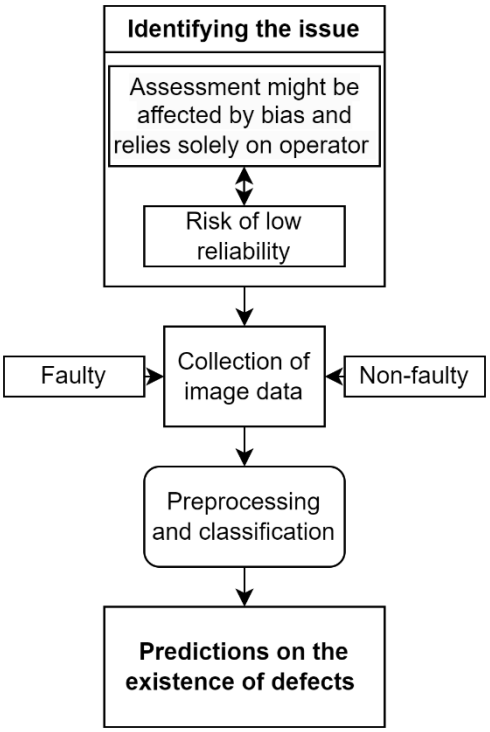


Figure 2: The approach to designing the data-driven method.

According to several researchers (Ali et al., 2012; Bertovic et al, 2013), the examination of a part requires information processing that contains signal detection and decision-making. The first decision-making at this point is not to identify the specific type of defect but to determine whether the picture of the cut material contains a defect or not. In signal detection, the aim is to recognize a signal from a background interference or noise (Swets, 1996). Therefore, the operator can give two right or wrong answers: to correctly or incorrectly accept or reject the presence of a defect (Enkvist et al., 1999; Lynn &Barrett,2014). The four possible outcomes are illustrated in Figure 3.

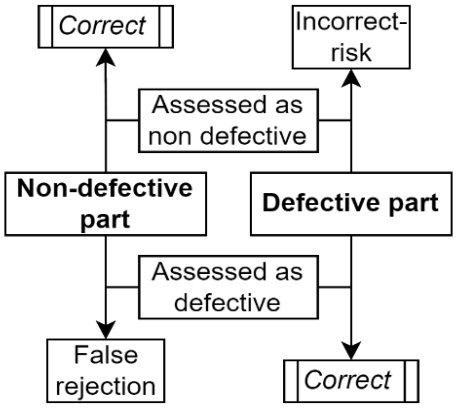


Figure 3: The possible outcomes of the assessment depend on the true state of the world, according to signal detection theory.

2.2. Data Preprocessing

The input for the model was both images that contained defects and images that did not. The images in the dataset were taken from a database of microscopic examinations and were previously used to manually inspect portions of the parts to find defects. Initially, 1787 photographs were retrieved from the database that had various kinds of defects, while 462 images had no signs of any defects. Without a form of data augmentation that would provide a wider and more balanced training dataset, it can be challenging to obtain appropriate performance because datasets from real applications (such as production) are frequently limited (Shorten & Khoshgoftaar, 2019; Xu et al., 2023).

For the initial processing of the images, data augmentation was applied. A usual form of data augmentation technique is altering the geometrical characteristics of the initial images.

The first step of augmentation consisted of using the PIL Python library to alter the dimensions of the images. PIL is widely used in Python, as a potent tool for processing images. It can alter different kinds of image formats, sizes, and orientations (Guan et al., 2019). The images were cropped to the ratio of 1:1, to facilitate the rotation process that took place later. The initial dimensions were 722*990 pixels, and the final images were 722*722 pixels, cropped regarding the defect area. To increase the number of training data, the images were flipped and subjected to rotation.

Figure 4 displays the original and altered photos as well as instances of the four categories of defects that were stated before.

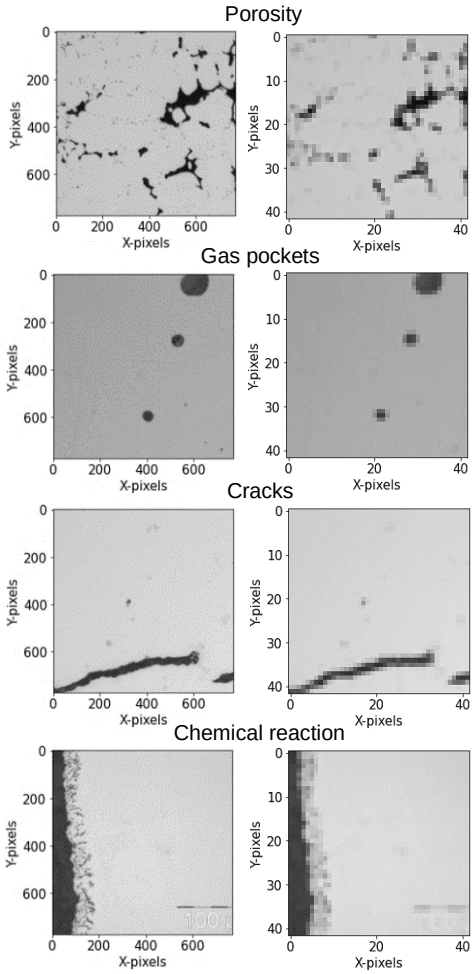


Figure 4: The original cropped images (left) of and the artificially edited ones (right) of four different types of defects (from top to bottom): porosity, gas pockets, cracks, and chemical reaction.

To enhance the computational performance of the classifier, the image pixels were reduced to 42*42 instead of 722*722. The classifier was able to better predict the existence of a defect on the image tested when the pixels were reduced to 42 per side, as shown in Figure 5, which indicates the higher accuracy levels achieved with this particular number of pixels.

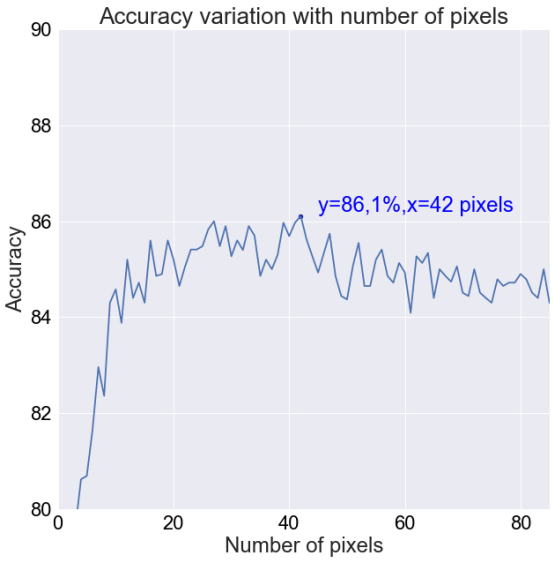


Figure 5: The accuracy score development during the repetitions with different numbers of pixels.

2.3. Classification

Considering the specific dataset's characteristics, including its size and the presence of defects, the proposed approach employs the ML classifier Random Forest (RF) to make predictions regarding the presence of defects in the input image. In applications with datasets similar to the one being utilized in this study (Khatami et al., 2019; Subudhi et al., 2020), RF shows satisfactory performance. This remains accurate for this stage of the process and is in line with the requirements of image recognition and classification between the two classes. Opting for RF over alternative methods is supported by its dependable performance in effectively tackling the challenges presented by dataset size and complexity. According to the literature, RF is often used for small data sets, similar to those from the medical field because it contributes to solving problems in industrial applications and has advantages such as ease of use, robust generalization ability, greater classification accuracy, and high functionality (Wang et al., 2023).

The RF method is considered quite a popular ensemble technique for pattern and image recognition. As an ensemble learning technique, it combines multiple decision trees to increase the

predictions' accuracy. The training is accomplished for each decision tree, where all classifiers generated from different trials are collected to construct the final classifier (Azar et al., 2014). The algorithm, when used for classification, outputs the mode of the classes of the individual trees. A subset of training data and a subset of features are randomly selected by the algorithm to build each decision tree.

The limitations of an RF classifier would depend on the high dimensionality of the data, which was tackled by reducing the number of pixels during data preprocessing, as mentioned in section 2.2. For pixel-based approaches like the one in this application, and for this stage and the requirements of the process, RF can perform satisfactorily. Research on methods for pixel analysis of pictures has revealed that RF acts similar to Neural Networks (NN) in defining linear borders between classes, such as in the usage of plantation boundaries (Boston et al., 2022). When aiming to reduce time consumption and computational complexity while dealing with a small number of training samples, RF has been preferred over NN in situations with diagnostic applications comparable to this one (Han et al., 2018).

The images obtained from previous microscope examinations were pre-processed and split into training and testing datasets. One of the possible limitations in training the RF classifier would be an imbalanced data set, where the classifier might favor the minority class. Therefore, a balanced data of 3600 images from each category (faulty or non-faulty) was used to create the training and testing datasets, to better assist the training process and reduce computational complications. The data set split was 80% training and 20% testing images, which was 5760 and 1440 images respectively.

Since the data set was labeled during the preprocessing stage, comparable supervised machine learning classifiers have been employed on this dataset. These included the Decision Tree (DT), the Support Vector Machine (SVM), and the Gaussian Process (GP) Classifier. The goal of the DT classifier is to create a training model that can be used to infer learning decision rules from training data in order to predict the class or value of target variables (Charbuty & Abdulazeez, 2021). The SVM is a common pattern recognition classification technique that aims to find a central hyperplane to partition the data points. The datasets are therefore divided into different classes. Along the hyperplane that separates the classes, SVM establishes a concentrated separation boundary (Halder et al., 2023). GP classifiers offer a probability distribution over all conceivable functions that can match a given set of training points. The decision boundary then corresponds to the midpoint between the two classes as a result of the prior distribution's initial assignment of equal probability to both classes (Basha et al., 2023). On the basis of their accuracy score, the three aforementioned techniques were compared with the RF classifier.

3. Results and discussion

The RF classifier underwent testing with different numbers of estimators to determine the best configuration that would produce the most accurate outcomes. It attained an accuracy rate of 86.5%. This score was found to be higher after experimenting with several types of classifiers.

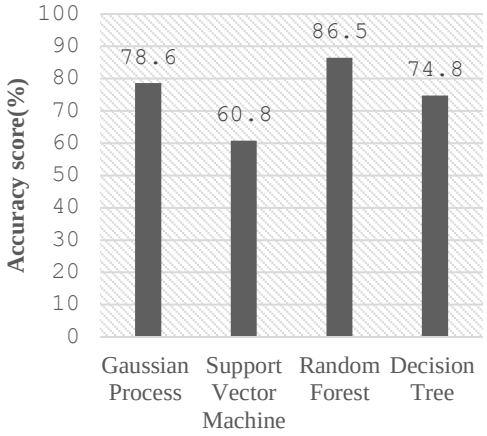


Figure 6: The accuracy score between the different classifiers.

As illustrated in Figure 6, the RF classifier outperformed other classifiers used in comparable applications, for this particular dataset with industrial images. The other types of classifiers that were tested and produced accuracy scores were the GP (78,6%), the SVM (60,8%), and finally the DT classifier (74,8%).

The number of assigned estimators, which in this application was 120 estimators, is typically used to describe the Random Forest classifier. This was obtained by several iterations of the model, each using a different set of estimators. After achieving peak accuracy at the 120 estimators (86.5%), it was seen that the computing time increased while the accuracy score did not, entering a relatively static period.

The development of the accuracy score regarding the number of estimators used is illustrated in Figure 7.

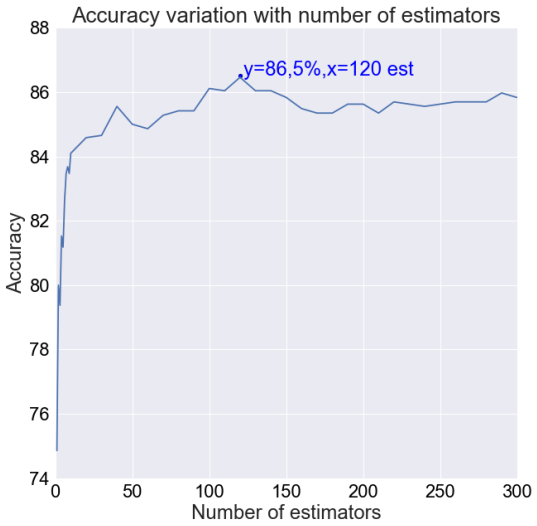


Figure 7: The accuracy score development during the repetitions with different estimator numbers.

The model produced a confusion matrix, a metric used to evaluate the accuracy of the classification. The confusion matrix contrasted the amount of accurate and inaccurate classifications made. Correct forecasts outperformed incorrect ones (false calls or misses) by a factor of five in terms of outcomes. The color gradient scale of the confusion matrix draws attention to the accurate classifications and the stark contrast between them and the inaccurate ones. The three matrices for the other classifiers were produced in addition to the confusion matrix from the RF classifier (Figure 8).

As observed in the confusion matrices for each classifier, the RF classifier demonstrates a more even distribution along the diagonal of the color scale. It achieves 1240 correct predictions (composed of 642 true positives and 548 true negatives), as opposed to 200 incorrect predictions (comprising 78 false positives and 122 false negatives). This pattern aligns with the accuracy scores, as the other classifiers show a decreased frequency of accurate predictions that match the true labels. The color-coded cells within the matrices distinctly indicate that only the SVM classifier surpasses the RF classifier in prediction count for a specific class. However, the SVM's incorrect predictions outnumber the correct ones, resulting in a lower accuracy score for this classifier. In essence, the analysis underscores that the RF classifier outperforms the others by maintaining a more balanced and accurate distribution of predictions, making it the most reliable choice among the evaluated classifiers.

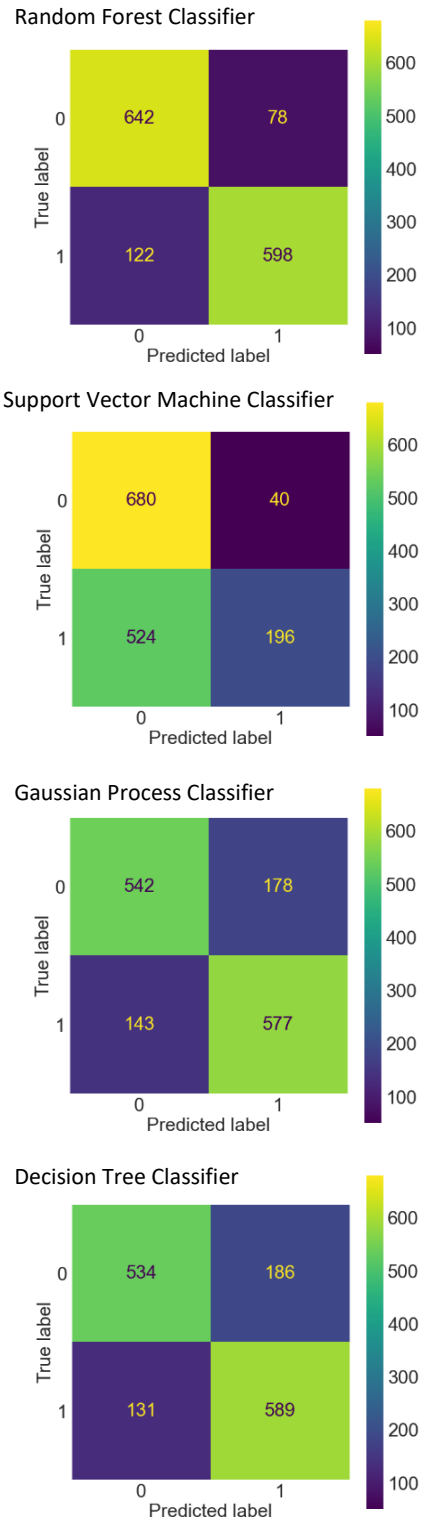


Figure 8: The Confusion Matrix for each classifier.

4. Conclusions

The present work demonstrated a data-driven approach to the investment casting microscopic examination to provide an assisting tool that supports decision-making and improves its reliability. The RF classifier that was chosen has achieved a level of prediction accuracy that is adequate given the characteristics of the dataset that was collected and preprocessed. It established higher efficiency for the selected dataset when compared to other classifiers that are employed in similar applications. Even though production data are seldom balanced, the model may also be used to predict unbalanced datasets after it has been trained. Therefore, this work contributes to developing a framework for integrating machine learning into the investment casting process, particularly in one of its subprocesses. It encourages further use of the ML classification algorithms for investment casting defects while introducing semi-automation of the investment casting microscopic examination.

Acknowledgments

The authors would like to thank TPC Components AB. The company's data contribution and support in the development of the data-driven technique are much appreciated. Many thanks also to Hassan Fazal and Sara Saketi, former quality engineers at TPC, for their invaluable assistance and knowledge sharing during the data retrieval and preparation process.

References

Ali, A. H., Balint, D., Temple, A., & Leevers, P. (2012). The reliability of defect sentencing in manual ultrasonic inspection. *NDT and E International*, *51*, 101–110. https://doi.org/10.1016/j.ndteint.2012.04.003

Aust, J., Mitrovic, A., & Pons, D. (2021). Assessment of the effect of cleanliness on the visual inspection of aircraft engine blades: An eye tracking study. *Sensors*, *21*(18). https://doi.org/10.3390/s21186135

Azar, A. T., Elshazly, H. I., Hassanien, A. E., & Elkorany, A. M. (2014). A random forest classifier for lymph diseases. *Computer Methods and Programs in Biomedicine*, *113*(2), 465–473. https://doi.org/10.1016/j.cmpb.2013.11.004

Basha, N., Kravaris, C., Nounou, H., & Nounou, M. (2023). Bayesian-optimized Gaussian process-based fault classification in industrial processes. *Computers and Chemical Engineering*, *170*. https://doi.org/10.1016/j.compchemeng.2022.108126

Boston, T., Van Dijk, A., Larraondo, P. R., & Thackway, R. (2022). Comparing CNNs and Random Forests for Landsat Image Segmentation Trained on a Large Proxy Land Cover Dataset. *Remote Sensing*, *14*(14). https://doi.org/10.3390/rs14143396

Charbuty, B., & Abdulazeez, A. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, *2*(01), 20–28. https://doi.org/10.38094/jastt20165

Del Vecchio, C., Fenu, G., Pellegrino, F. A., Michele, D. F., Quatrale, M., Benincasa, L., Iannuzzi, S., Acernese, A., Correra, P., & Glielmo, L. (2019). Support Vector Representation Machine for superalloy investment casting optimization. *Applied Mathematical Modelling*, *72*, 324–336. https://doi.org/10.1016/j.apm.2019.02.033

Dezecot, S., & Brochu, M. (2015). Microstructural characterization and high cycle fatigue behavior of investment cast A357 aluminum alloy. *International Journal of Fatigue*, *77*, 154–159. https://doi.org/10.1016/j.ijfatigue.2015.03.004

Enkvist, J., Edland, A., & Svenson, O. (1999). *Human Factors Aspects of Non-Destructive Testing in the Nuclear Power Context A review of Research in the Field*.

Esmaeilian, B., Behdad, S., & Wang, B. (2016). The evolution and future of manufacturing: A review. In *Journal of Manufacturing Systems* (Vol. 39, pp. 79–100). Elsevier B.V. https://doi.org/10.1016/j.jmsy.2016.03.001

García Pérez, A., Gómez Silva, M. J., & de la Escalera Hueso, A. (2022). Automated Defect Recognition of Castings Defects Using Neural Networks. *Journal of Nondestructive Evaluation*, *41*(1). https://doi.org/10.1007/s10921-021-00842-1

González Rodríguez, G., Gonzalez-Cava, J. M., & Méndez Pérez, J. A. (2020). An intelligent decision support system for production planning based on machine learning. *Journal of Intelligent Manufacturing*, *31*(5), 1257–1273. https://doi.org/10.1007/s10845-019-01510-y

Guan, Y., Zhou, F., & Zhou, J. (2019). Research and Practice of Image Processing Based on Python. *Journal of Physics: Conference Series*, *1345*(2). https://doi.org/10.1088/1742-6596/1345/2/022018

Halder, S., Das, S., & Basu, S. (2023). Use of support vector machine and cellular automata methods to evaluate impact of irrigation project on LULC. *Environmental Monitoring and Assessment*, *195*(1). https://doi.org/10.1007/s10661-022-10588-6

Han, T., Jiang, D., Zhao, Q., Wang, L., & Yin, K. (2018). Comparison of random forest, artificial neural networks and support vector machine for intelligent diagnosis of rotating machinery. *Transactions of the Institute of Measurement and Control*, *40*(8), 2681–2693. https://doi.org/10.1177/0142331217708242

Hao, Y., Liu, J., Du, J., Zhang, W., Xiao, Y., Zhang, S., & Yang, P. (2020). Effects of mold materials on the interfacial reaction between magnesium alloy and ceramic shell mold during investment casting. *Metals*, *10*(8), 1–14. https://doi.org/10.3390/met10080991

Jawahar, M., Babu, N. K. C., Vani, K., Anbarasi, L. J., & Geetha, S. (2021). Vision based inspection system for leather surface defect detection using fast convergence particle swarm optimization ensemble classifier approach. *Multimedia Tools and Applications*, *80*(3), 4203–4235. https://doi.org/10.1007/s11042-020-09727-3

Jiang, L., Wang, Y., Tang, Z., Miao, Y., & Chen, S. (2021). Casting defect detection in X-ray images using convolutional neural networks and attention-guided data augmentation. *Measurement: Journal of the International Measurement Confederation*, *170*. https://doi.org/10.1016/j.measurement.2020.108736

John A. Swets. (1996). *Book Signal Detection Theory and ROC Analysis in Psychology and Diagnostics* (1st Edition). eBook Published 21 January 2014.

Kaiser, R., Browne, D. J., & Williamson, K. (2011). Investigation of the effects of cooling rate on the microstructure of investment cast biomedical grade Co alloys. *IOP Conference Series: Materials Science and Engineering*, *27*(1). https://doi.org/10.1088/1757-899X/27/1/012071

Khatami, A., Araghi, S., & Babaei, T. (2019). Evaluating the performance of different classification methods on medical X-ray images. *SN Applied Sciences*, *1*(10). https://doi.org/10.1007/s42452-019-1174-0

Li, F., & Wang, F. (2021). Preparation Process of Ceramic Shells. In *Precision Forming Technology of Large Superalloy Castings for Aircraft Engines* (pp. 101–167). Springer Singapore. https://doi.org/10.1007/978-981-33-6220-8_4

Lynn, S. K., & Barrett, L. F. (2014). "Utilizing" Signal Detection Theory. *Psychological Science*, *25*(9), 1663–1673. https://doi.org/10.1177/0956797614541991

MacLean, C. L., & Dror, I. E. (2021). The effect of contextual information on professional judgment: Reliability and biasability of expert workplace safety inspectors. *Journal of Safety Research*, *77*, 13–22. https://doi.org/10.1016/j.jsr.2021.01.002

Marija Bertovic, phil. (2013). Human Factors in Non-Destructive Testing (NDT): Risks and Challenges of Mechanised NDT. *Materials Testing*. https://doi.org/10.14279/depositonce-4685

Mohamed, M., & Saber, N. El. (2023). Intelligent Decision Support Machine Learning Based Optimizing Inventory Management. *American Journal of Business and Operations Research*, *9*(2), 41–50. https://doi.org/10.54216/AJBOR.090205

Mozammil, S., Verma, R., Karloopia, J., & Jha, P. K. (2020). Investigation and measurement of porosity in Al + 4.5Cu/6wt%TiB2 in situ composite: optimization and statistical modelling. *Journal of Materials Research and Technology*, *9*(4), 8041–8057. https://doi.org/10.1016/j.jmrt.2020.05.045

Niccolai, A., Caputo, D., Chieco, L., Grimaccia, F., & Mussetta, M. (2021). Machine learning-based detection technique for ndt in industrial manufacturing. *Mathematics*, *9*(11). https://doi.org/10.3390/math9111251

Safhi, H. M., Frikh, B., & Ouhbi, B. (2019). Assessing reliability of Big Data Knowledge Discovery process. *Procedia Computer Science*, *148*, 30–36. https://doi.org/10.1016/j.procs.2019.01.005

Shipway, N. J., Barden, T. J., Huthwaite, P., & Lowe, M. J. S. (2019). Automated defect detection for Fluorescent Penetrant Inspection using Random Forest. *NDT and E International*, *101*, 113–123. https://doi.org/10.1016/j.ndteint.2018.10.008

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, *6*(1). https://doi.org/10.1186/s40537-019-0197-0

Subudhi, A., Dash, M., & Sabut, S. (2020). Automated segmentation and classification of brain stroke using expectation-maximization and random forest classifier. *Biocybernetics and Biomedical Engineering*, *40*(1), 277–289. https://doi.org/10.1016/j.bbe.2019.04.004

Torroba, A. J., Koeser, O., Calba, L., Maestro, L., Carreño-Morelli, E., Rahimian, M., Milenkovic, S., Sabirov, I., & LLorca, J. (2014). Investment casting of nozzle guide vanes from nickel-based superalloys: part I – thermal calibration and porosity prediction. *Integrating Materials and Manufacturing Innovation*, *3*(1), 344–368. https://doi.org/10.1186/s40192-014-0025-5

Wang, J., Rao, C., Goh, M., & Xiao, X. (2023). Risk assessment of coronary heart disease based on cloud-random forest. *Artificial Intelligence Review*, *56*(1), 203–232. https://doi.org/10.1007/s10462-022-10170-z

Warren, P., Ali, H., Ebrahimi, H., & Ghosh, R. (2021). *RAPID DEFECT DETECTION AND CLASSIFICATION IN IMAGES USING CONVOLUTIONAL NEURAL NETWORKS*.

Xu, M., Yoon, S., Fuentes, A., & Park, D. S. (2023). A Comprehensive Survey of Image Augmentation Techniques for Deep Learning. *Pattern Recognition*, *137*. https://doi.org/10.1016/j.patcog.2023.109347

Process Simulation, Dimensioning and Automated Cost Optimization of CO_2 Capture

Lars Erik Øi*, Shirvan Shirdel, Sumudu Karunarathne, Solomon Aromada

Department of Process, Energy and Environmental Technology, University of South-Eastern Norway
lars.oi@usn.no

Abstract

A standard process for CO_2 capture has been simulated with an equilibrium-based model in Aspen HYSYS. The simulation has been combined with equipment dimensioning and cost calculation in an integrated spreadsheet facility. New in this work is that Murphree efficiencies are varied to obtain automatic optimization of absorber height and inlet temperature. The optimum process was found as the process with minimum calculated sum of capital and operational cost over 25 years. The cost optimum process parameters for the standard process were calculated to 15 m absorber packing height, 13 K minimum approach temperature and 34 °C in inlet gas temperature. This study demonstrates that it is possible to calculate the optimum packing height and inlet temperature automatically by varying the Murphree efficiency in a case study function.

Keywords: Carbon capture, Aspen HYSYS, simulation, cost estimation, optimization

1. Introduction

1.1. Aim

The general aim of this work is to calculate the cost optimum absorption column height, minimum temperature approach temperature in the main amine/amine heat exchanger and optimum inlet temperature to the absorber. A specific aim is to make it possible to calculate these optimums automatically by varying the Murphree efficiency.

1.2. Literature

Much work has been published on cost estimation of CO_2 capture plants (Rubin *et al*., 2013; van der Spek *et al*., 2019; Roussanaly *et al*., 2021). Several papers present results from process simulation and cost estimation (Mores *et al.*, 2012; Agbonghae *et al.*, 2014); Manzolini *et al*., 2015; Luo and Wang, 2016; Nwaoha *et al*., 2018; Eldrup *et al*., 2019; Hasan *et al*., 2021).

Some of the previous works at Telemark University College and the University of South-Eastern Norway (USN) with focus on process simulation, equipment dimensioning, cost estimation and optimization are Kallevik (2010), Øi (2007), Øi (2012), Aromada and Øi (2017) and Øi *et al*. (2022). The cost estimation part has in most of these works been based on different detailed factor methods like the Enhanced Detailed Factor (EDF) method (Ali *et al*., 2019; Aromada *et al*., 2021). In these works, the main approach for calculating the optimum has been to use case studies in Aspen HYSYS and varying only one parameter at a time. Then the optimum is found as the simulation giving the minimum sum of capital and operational cost.

In the recent years, a focus has been on automatic process simulation combined with cost estimation in Aspen HYSYS (Øi *et al*., 2021; Øi *et al*., 2022; Shirdel *et al.*, 2022). An Iterative Detailed Factor (IDF) scheme was developed (Aromada *et al*., 2022a) where an aim was to make the entire process simulation, equipment dimensioning and cost estimation automatic, without requiring any manual input. This was accomplished in the work by Øi *et al*. (2022) by linking Aspen HYSYS simulation spreadsheets with Microsoft Excel by a VBA (Visual Basic) code. With an automated approach, process simulation based CO_2 capture, process parameter cost optimization studies and sensitivity analysis can be conducted quickly and obtain reasonably accurate results.

A limiting factor for automation in the Aspen HYSYS tool, has been that for a column, the number of equilibrium stages must be changed manually. To overcome this, a possibility is to vary the Murphree efficiency on one or a selected number of absorption stages. The optimization can then be performed by performing a case study in Aspen HYSYS. This work is based on the results from the Master thesis work of Shirdel (2022), and in addition more references are included and discussed.

1.3. Process Description

The use of an amine solvent to remove CO_2 is the most widely used and well-studied approach for CO_2 removal. Monoethanol amine (MEA) is the solvent that has been studied most, and it works well due to its quick interaction with CO_2. Fig. 1 is a typical process flow diagram for an amine-based CO_2 removal facility. Traditional absorption is done in a column using plates, random packing or structured packing. The CO_2-containing gas rises, while the absorption liquid falls. The solvent (rich amine) is then fed to a desorption column through a heat exchanger. In the desorption (stripper) column, the CO_2 that has been absorbed is regenerated. The reboiler is heated, and a condenser provides reflux to the column. The regenerated solvent (lean amine) is recirculated to the absorption column after the desorber and cooled in a heat exchanger and cooler.

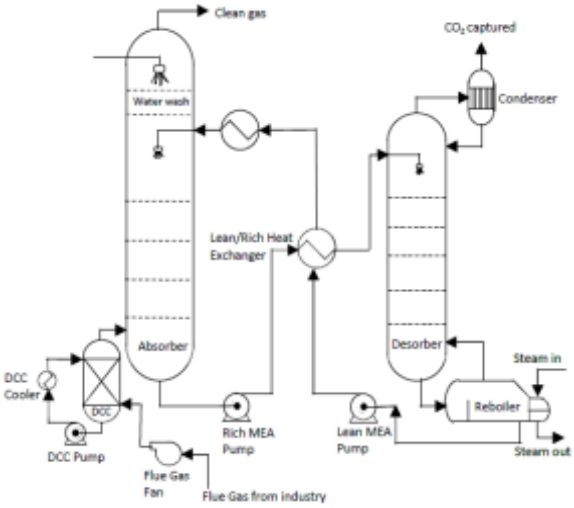


Figure 1: Process flow diagram of a standard amine-based CO_2 capture process (Aromada *et al*., 2020).

2. Specifications and simulations

2.1. Specifications and simulation of base case CO_2 capture process

In this investigation, Aspen HYSYS version 12 was used to model a conventional amine-based CO_2 capture process, and the simulated results were utilized to size equipment and estimate costs using the same calculation method as in Aromada et al. (2021), Øi *et al.* (2021) and Øi *et al.* (2022). In all simulations, the Acid Gas property package was employed, which includes a liquid equilibrium model for electrolytes. This package is intended to replace the Amine property package, which has been widely used when using the Aspen HYSYS tool. The electrolyte non-random two-liquid (e-NRTL) model for electrolyte thermodynamics and the Peng-Robinson equation of state for the vapor phase were used to create this property package.

The absorber and desorber were simulated using equilibrium stages containing user defined stage (Murphree) efficiencies. These Murphree efficiencies are defined by dividing the change in CO_2 mole fraction from one stage to the next by the change on the assumption of equilibrium.

Emission data from previous studies (Aromada and Øi, 2017) on a natural gas-based power plant project on Mongstad, Norway, were utilized to generate the base case for the simulations. The specifications in Table 1 correspond to an 85 per cent CO_2 removal efficiency and a minimum approach temperature of 10 °C in the lean/rich amine heat exchanger, which is considered the base case configuration. 85 % CO_2 removal rate is traditional for capture from power plant based on natural gas. The absorber is modelled with 15 packing stages, while the desorber has 10. Murphree efficiencies of 15% were employed in the absorption column and 50% for all stages of the desorption column where one stage is expected to be approximately 1 meter packing height. In the columns, the Modified HYSIM Inside-Out numerical solver was adopted since it assists in convergence. The adiabatic efficiency of the pump and flue gas fan was specified to be 75%.

To obtain an automated simulation model, robust adjustments and recycles are necessary to aid in the convergence of the simulations. Traditionally, manual adjustments can be performed by trial and error when working with a complex simulation model.

The calculation sequence is similar to the simulations in Øi *et al.* (2021) and Øi *et al.* (2022). It starts with the input gas and the lean amine to the absorption column (which is first guessed). The rich amine pump transports the rich amine from the bottom of the absorption column through the lean/rich amine heat exchanger. After the heat exchanger, the temperature is specified, and the rich amine is sent to the desorber. The CO_2 product and the hot lean amine are calculated in the desorption column. The heated lean amine is passed via the lean/rich heat exchanger and then pressurized in the lean amine pump, before being cooled further in the lean cooler. Water was added to the process (water make-up) and the make-up was calculated by a water material balance.

The lean amine is then placed in a recycle block (RCY_1). It is determined whether the recycled lean amine's flow and condition are sufficiently similar to the previously estimated lean amine stream, which may be adjusted through iteration.

In order to create an automated simulation model, three adjust operations were added to the flowsheet. The removal efficiency can be adjusted based on the lean amine flow rate by ADJ-1, the minimum approach temperature in the lean/rich heat exchanger may be adjusted based on the rich amine outlet temperature of the lean/rich heat exchanger by

ADJ-2, and for adjusting the flue gas temperature to the absorber, ADJ-3 changes the cooling water supply in the inlet cooler. The default tolerances in Aspen HYSYS were used in the simulations.

Table 1: Specifications for the base case alternative

Parameter	Value
Inlet flue gas temperature [°C]	80/40.0
Inlet flue gas pressure [kPa]	101/115
Inlet flue gas flow rate [kmol/h]	85000
CO_2 content in inlet gas [mole %]	3.75
Water content in inlet gas [mole %]	6.71
Lean amine temperature [°C]	40.0
Lean amine pressure [kPa]	101.0
Lean amine rate [kg/h]	103500
MEA content in lean amine [mass %]	29
CO_2 content in lean amine [mass %]	5.5
Number of stages in absorber [-]	15
Murphree efficiency in absorber	0.15
Rich amine pump pressure [kPa]	200.0
Rich amine temp. out of HEX [°C]	103.7
Number of stages in desorber [-]	10
Murphree efficiency in desorber	0.5
Reflux ratio in stripper [-]	0.3
Reboiler temperature [°C]	120.0
Lean amine pump pressure [kPa]	500.0

2.2. Parameter variation of ΔT_{min}

A case study was made to look into the economic performance of the lean/rich amine heat exchanger when the degree of heat recovery was adjusted. More heat recovery will normally increase the capital cost and reduce the operating cost. The ΔT_{min} was changed for each simulation case. This was performed automatically by changing the target temperature from 5 to 20 °C in ADJ-2, whereas the ADJ-1 and ADJ-3 will aim to maintain a constant CO_2 removal efficiency of 85% and a constant incoming flue gas temperature of 40 °C, respectively. All flue gas and absorption column parameters were held constant throughout the case for a certain total CO_2 removal efficiency, lean amine composition and lean amine flow.

2.3. Parameter variation of number of absorption stages and absorber height

A higher absorption column packing is expected to increase the capital cost and reduce the operating cost. Because each change in the number of stages in the Design tab of the absorber requires manual input to run the simulation again, the case study option cannot be utilized in the sensitivity analysis for altering absorber height (stages). In all stages, Murphree efficiency has been set to 0.15. For each case, the efficiency of new stages, the pressure of flue gas into the absorber, and the pressure in the absorber's last stage should all be updated. For this reason, a new spreadsheet was created, and the calculations for changing the number of stages of the absorber and fan outlet pressures based on 1 kP for each stage (Park and Øi, 2017) were performed.
In this study, a strategy was employed to define a case study by altering the efficiency of one specific stage. Changing the efficiency at one stage from 0.15 to 0.9 for a configuration with 13 stages is almost equivalent to increase the number of stages from 13 to 18. Throughout the case study, the absorber efficiency, lean/rich amine heat exchanger minimum temperature approach, all flue gas parameters and lean amine content were all kept constant. The lean amine feed in ADJ-1, the desorber input temperature in ADJ-2, the flow rate of inlet cooling water in the inlet heat exchanger in ADJ-3 and the mass balance of makeup MEA and water in the makeup streams spreadsheet had to be adjusted to maintain the specification values.

2.4. Parameter variation of absorber inlet temperature

An analysis was conducted to adjust the flue gas inlet temperature to the absorber column. A high column temperature will increase the absorption rate and reduce the CO_2 solubility, so it is expected that the inlet temperature has an optimum. This is done in ADJ-3 by altering the cooling water input flow rate and as a result also changing the absorber inlet temperature. The lean amine composition was kept constant (by defining the MakeUp Streams spreadsheet), but the lean amine flow rate was adjusted in ADJ-1 for each case to obtain the desired CO_2 removal efficiency. The ADJ-2 operates to keep the ΔT_{min} constant in the lean /rich amine heat exchanger.
It is possible to specify the Murphree efficiency for each absorber stage. The Murphree efficiency must be adjusted for each new inlet temperature, which makes this calculation complex. The Murphree stage efficiency was adjusted to account for the impacts of varying temperature profiles in the absorber column at various input gas temperatures. Øi (2012) has made a computational approach for estimating the Murphree stage efficiency as a function of temperature for absorber top and bottom conditions. Based on this calculation scheme, the Murphree efficiencies were computed only for the top-, bottom-, and maximum temperature stages, and the intermediate stage temperatures have been obtained

using a linearization between these temperatures as done in Kallevik (2010).
After calculating the average Murphree efficiency for each inlet flue gas temperature, an equation for the relationship between the inlet temperature and Murphree efficiency was made. Another spreadsheet was created to export the calculated stages efficiency to the absorber after changing the incoming flue gas temperature from 30 to 50 °C in the case study.

3. Cost estimation procedures and assumptions

3.1. Equipment dimensioning and assumptions

The mass and energy balances from the Aspen HYSYS process simulations were used to dimension all the equipment as done in previous studies (Øi *et al.*, 2022).

The diameters of the absorption and desorption columns were evaluated from the gas stream's volumetric flows. These were based on superficial gas velocities of 2.5 m/s and 1 m/s respectively as done in earlier studies (Aromada & Øi, 2017). In the base case, 15 packing stages were specified for absorber, and 10 for the desorber. Each packing stage in the absorber and desorber were assumed to be 1 m high (Aromada *et al*., 2020). Structured packing was specified for better operational cost due to pressure drop (Choi *et al*., 2005). To estimate the tangent-to-tangent height of the absorber, the packing, liquid distributors, water wash, demister, gas inflow and outflow and sump were all considered. The condenser inlet, packing, liquid distributor, gas input and sump were taken into account in estimation of the desorber tangent-to-tangent height (Ali, 2019; Øi *et al*., 2021). The packing height was given from a design of a wash tower in the catalog (Sulzer Chemtech, 2021). Thus, 35 m and 25 m were arrived at for the tangent-to-tangent heights of absorber and desorber respectively.

The separator was sized using Souders Brown's equation with a k-factor of 0.15 m/s and a height to diameter ratio of 1. The heat duties obtained from the process simulations were used to size the heat exchange equipment. The overall heat transfer coefficients specified are 1.20 kW/(m²·K) for the reboiler, 0.73 kW/(m²·K) for the lean/rich heat exchanger, 0.80 kW/(m²·K) for the amine cooler, and 1.00 kW/(m²·K) for the condenser as in Aromada *et al*. (2022b). The pumps, compressor and fan were sized based on their duties with efficiency 75 %.

3.2. Capital cost estimation method

The Enhanced Detailed Factor (EDF) method (Ali *et al*., 2019; Aromada *et al*., 2021) was applied for the estimation of the CO_2 capture plant's capital cost. As a detailed factor approach, the installed cost of each equipment is estimated based on variable installation factors that depends on each equipment cost. The capital cost of the CO_2 capture plant is then the sum of all equipment installed costs. The updated EDF factor list is published in (Aromada *et al*., 2021).

Each equipment unit delivered cost was obtained from Aspen In-plant Cost Estimator (v.12). This is based on the capacity or size of each of the equipment units as determined from the dimensioning process. The cost currency and cost year were Euro (€) and 2019. The default location in Aspen In-Plant Cost Estimator, Rotterdam, was assumed in this work. The equipment units were assumed to be constructed from stainless steel (SS316). To apply the EDF method, the cost of the equipment units must be converted from their costs in the original material of construction. The cost (EC_{SS}) of an equipment unit in stainless steel (SS) needs to be converted to its cost (EC_{CS}) in carbon steel (CS). This is implemented by applying a material factor (f_{mat}) where CS is the reference material. The cost of each equipment unit constructed in welded SS is divided by a material factor of 1.75 to convert it to the corresponding cost in CS material. While the material factor for units manufactured in machined SS, e.g. pumps, is 1.30. Then, the total installation factor ($F_{T,CS}$) and piping subfactor (f_{pp}) in CS for each equipment unit are obtained from the EDF factor lists (Aromada *et al.*, 2021). They are then converted to total installation factor in SS ($F_{T,SS}$) as shown in equation (1):

$$F_{T,SS} = \left[F_{T,CS} + \left\{(f_{mat} - 1)\left(f_{eq} + f_{pp}\right)\right\}\right] \quad (1)$$

Where f_{eq}= equipment factor = 1.0

The total equipment installed cost (*EIC*) is estimated as follows:

$$EIC_{SS} = F_{T,SS} * EC_{CS} * (No.of\ units) \quad (2)$$

Then the total installed cost (CAPEX) with cost year of 2019 is:

$$CAPEX = \sum(EIC_{SS}\ for\ all\ equipment) \quad (3)$$

The capital cost of the CO_2 capture plant is then escalated from 2019 to 2021 using a consumer cost index from Statistisk Sentralbyrå (SSB). A Norwegian cost index is selected because the detailed factors were originally based on Norwegian currency.

During optimization or sensitivity analysis, where a parameter is varied, the capacities/sizes of some equipment will change. Therefore, there is a need to estimate new delivered cost for the equipment units due to the resulting changes in size/capacity. This is automatically estimated based on the Power law using an exponent of typically 0.65, from the previous cost obtained from Aspen In-Plant Cost

Estimator database as done in (Aromada *et al.*, 2022a, Aromada *et al.*, 2022b; Øi *et al.*, 2022).

3.3. Operating cost estimation and assumptions

The annual operating cost in this work is the sum of the fixed operating cost and variable operating costs. The variable operating cost was estimated from equation (4):

$$Annual\ variable\ cost\left(\frac{€}{yr}\right) = Consumption\ \left(\frac{unit}{hr}\right) \times \frac{Operating\ hours}{year} \times unit\ cost\ \left(\frac{€}{unit}\right) \quad (4)$$

The assumptions used for estimating the annual operating cost are presented in Table 2. The values are similar to values used in earlier work like Aromada et al. (2021). The steam cost is set to 25 % of the electricity cost because steam can be converted to electricity with an efficiency of order of magnitude 25 %.

Table 2: Annual operating cost assumptions.

Item	Unit	Value
Operating lifetime	[Year]	25[1]
Annual hours of operation	[h/year]	8000
Electricity cost	[€/kWh]	0.06
Steam cost	[€/kWh]	0.015
Cooling water cost	[€/m^3]	0.022
Water process cost	[€/m^3]	0.203
MEA cost	[€/ton]	1450
Maintenance cost	[€/year]	4% of CAPEX
Operator cost (6 oper)	[€/year]	80414(*6)
Engineer cost (1 eng)	[€/year]	156650

[1] 2 years construction + 23 years operation

3.4. CO_2 capture annualized cost

The economic key performance indicator in this work is CO_2 captured cost. This was estimated as:

$$CO_2\ captured\ cost = \frac{Total\ annual\ cost}{Mss\ of\ Captured\ CO_2/year} \quad (5)$$

$$Total\ annual\ cost = Annualized\ CAPEX + Yearly\ OPEX \quad (6)$$

$$Annualized\ CAPEX = \frac{CAPEX}{Annualized\ factor}$$

$$Annualized\ factor = \sum_{i=1}^{n}\left[\frac{1}{(1+r)^n}\right] \quad (7)$$

Where n is the plant lifetime, 25 years which includes 2 years for the plant's construction. And r is the discount rate and was assumed to be 7.5 %.

4. Results and Discussion

4.1. Base case cost results

The overall equipment cost was calculated to 110 MEUR, and the absorber is the costliest equipment, accounting for 54% of the total cost. This is traditional in other calculations (Ali 2019; Aromada *et al.*, 2021). The structured packing cost accounts for 55 percent of the absorber's total cost.

The total operational expenditure (OPEX) for the Base case was calculated to 29 MEUR/yr. Steam is the costliest utility for this facility, costing 15 MEUR each year. The steam usage is calculated to 3.75 GJ/ton CO_2 captured and this is in line with values in literature (Choi et al., 2005; Øi, 2012).

4.2. Optimization of minimum ΔT approach

CO_2 captured cost and energy consumption as a function of ΔT_{min} is shown in Fig. 3. It shows a flat minimum between 11 and 15, and a minimum at 13 K. Fig. 3 is based on an automated case study in Aspen HYSYS. The simulations were also calculated manually, obtaining a smoother curve because all the parameters could be adjusted more accurately by trial and error. The results were similar, but the optimum ΔT_{min} was calculated manually to 12 K. Similar values have been calculated in several works (Øi, 2012; Shirdel *et al.*, 2022). In the case of using plate heat exchangers, the optimum ΔT_{min} will be less than 10 K.

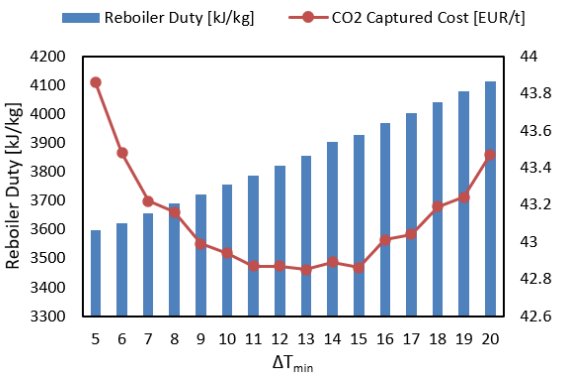


Figure 3: CO_2 captured cost and energy consumption as a function of ΔT_{min} (from Shirdel, 2022).

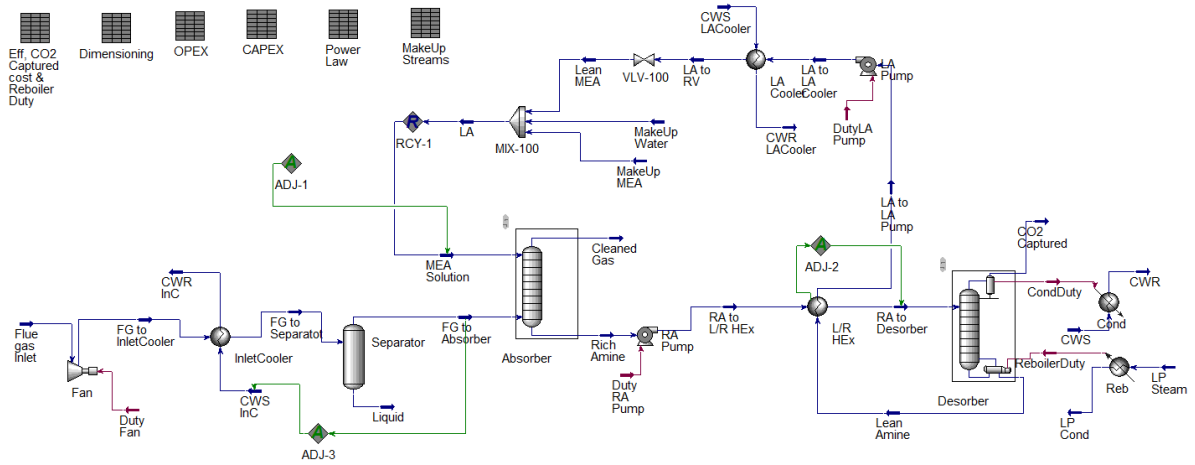


Figure 2: Aspen HYSYS flow sheet for the Base case simulation (from Shirdel, 2022).

4.3. Optimum absorber height

CO_2 captured cost and energy consumption as a function of absorber packing height is shown in Fig. 4. Results for both manual and automatic calculation are shown. Also here, the manual simulations give a smoother curve. However, the resulting optimum absorption height is 15 meters for both manual and automatic optimization. This is in the order of magnitude similar to earlier works where optimum packing height have been calculated to 20 meters (Mores *et al.*, 2012), 19 meters (Agbonghae *et al.*, 2014)) 15 meters (Aromada and Øi, 2017) and 19 meters (Shirdel *et al.*, 2022). All the heights were structured packing except for Mores et al. (2012) which was based on random packing.

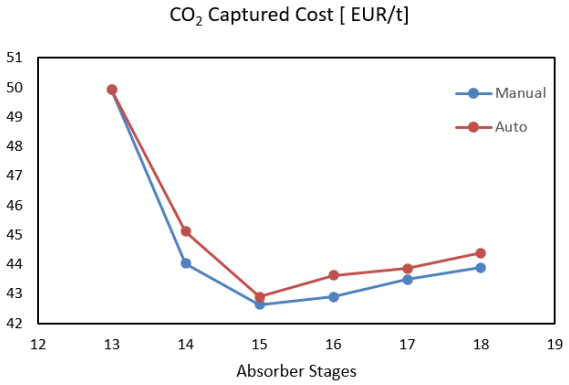


Figure 4: CO_2 capture cost as a function of absorber packing height (from Shirdel, 2022).

4.4. Optimum inlet gas temperature

To perform a reasonable optimization of the inlet gas temperature, the temperature dependence of the absorption efficiency must be included. In Fig. 6, the temperature and Murphree efficiency for the different absorption stages have been calculated. The Murphree efficiencies were calculated by the methods specified in Chapter 2, and one iteration was performed to include the effect of temperature on the calculated Murphree efficiencies from the first iteration.

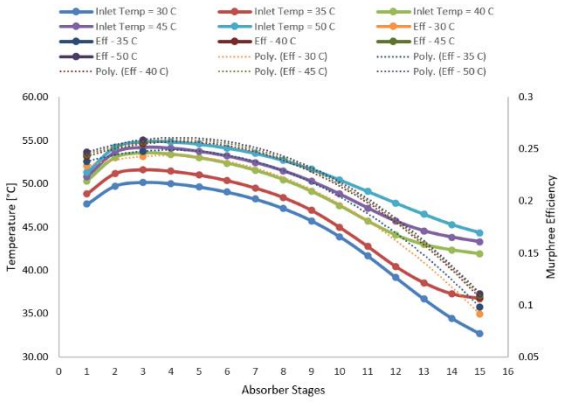


Figure 6: Murphree efficiency as a function of absorber stage and temperature (from Shirdel, 2022).

For each inlet gas temperature, an average Murphree efficiency was calculated by a fitted polynomial.

$$E_M = -0.00004T^2 + 0.0041\,T + 0.08 \qquad (8)$$

A preliminary optimization was performed by manual simulations of the CO_2 capture cost with 5 K steps for 15 and 13 absorption stages. The lowest cost case was found at 13 stages (meter of packing). The optimum was then calculated automatically in a case study for 13 absorption stages with temperature steps of 1 K in Fig. 7.

Fig. 7 shows that it is possible to calculate the optimum inlet gas temperature automatically. The curve is not very smooth, and this indicates that there are some inaccuracies in the calculations. To improve this, a possibility is to adjust the tolerances in the Aspen HYSYS simulation tool. This was evaluated by Øi et al. (2021). The most optimum point at the curve is for an inlet gas temperature of 34 °C. There are not found many numbers to compare with in literature, but Øi (2012) calculated an optimum between 33 and 35 °C.

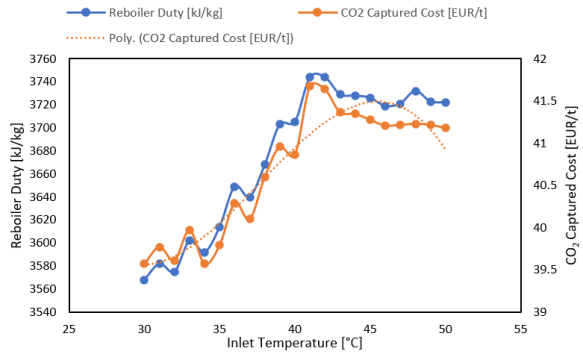


Figure 7: Optimization of inlet gas temperature (from Shirdel, 2022).

4.5. Optimization of other and several parameters

In this work, emphasis has been on the optimization of packing height, minimum temperature approach and inlet gas temperature. Other works have emphasized optimization of other parameters as absorber gas velocity and pressure drop (Park and Øi, 2017), CO_2 capture rate (Mores *et al.*, 2012) and lean loading (Agbonghae *et al.*, 2014). Optimization of these parameters are most often not independent. A high % CO_2 capture rate will e.g. give a lower optimum CO_2 loading.

Simultaneous optimization of several parameters have been evaluated by Mores *et al.* (2012) and Agbonghae *et al.* (2014). Mores *et al.* (2012) used a methodology based on Murphree efficiencies, and Agbonghae *et al.* (2014) based the work on rate-based modelling in Aspen Plus including the Aspen Plus Economic Analyser.

Such simultaneous optimization raises challenges for future work in complexity, accuracy, consistency and robustness of the calculations.

6. Conclusion

The case study function in Aspen HYSYS can be used to perform several simulations by changing one parameter at a time. The ΔT_{MIN} was optimum at 13 K (a flat optimum between 11-15 K) giving 42.8 EURO/ton CO_2. The case study function cannot be used to vary the number of stages in a column. However, the packing height was varied in an automated case study by increasing the Murphree efficiency of one stage gradually from 0.15 to 0.9. The optimum packing height at 15 meter (15 stages with 0.15 stage efficiency) gave 42.6 EURO/ton. Inlet temperature was optimized using the case study model where the Murphree efficiency was calculated as a function of temperature. Optimum inlet temperature was obtained at 34 °C, and the cost was reduced to 39.6 EURO/ton CO_2. The optimums agree well with earlier calculated optimum parameter values.

This study demonstrates that it is possible to calculate the optimum packing height and inlet temperature automatically by varying the Murphree efficiency in a case study function.

References

Ali, H. (2019) *Techno-economic analysis of CO_2 capture concepts.* PhD Thesis, University of South-Eastern Norway.

Agbonghae, E. O. et al. (2014), 'Optimal Process Design of Commercial-Scale Amine-Based CO_2 Capture Plants', *Industrial & Engineering Chemistry Research*, 53, pp. 14815-14829.

Aromada, S. A. and Øi, L.E. (2017) 'Energy and economic analysis of improved absorption configurations for CO_2 capture.' *Energy Procedia*, 114: pp. 1342-1351.

Aromada, S. A. *et al.* (2020) 'Simulation and Cost Optimization of different Heat Exchangers for CO_2 Capture', Linköping Electronic Conference Proceedings, SIMS 61, pp. 22-24. doi:10.3384/ecp20176318

Aromada, S. A. *et al.* (2021) 'Capital cost estimation of CO_2 capture plant using Enhanced Detailed Factor (EDF) method: Installation factors and plant construction characteristic factors' *International Journal of Greenhouse Gas Control*. 110, pp.103394.

Aromada, S. A. *et al.* (2022a) 'Simulation-based Cost Optimization tool for CO_2 Absorption processes: Iterative Detailed Factor (IDF) Scheme', *Scandinavian Simulation Society*, pp. 301-308.

Aromada, S. A. *et al.* (2022b) 'Cost and Emissions Reduction in CO_2 Capture Plant Dependent on Heat Exchanger Type and Different Process Configurations: Optimum Temperature Approach Analysis', *Energies*, 15(2), pp. 425.

G. N. Choi, R. Chu, B. Degen, H. Wen, P. L. Richen, and D. Chinn, CO_2 removal from power plant flue gas–cost efficient design and integration study, *Carbon Dioxide Capture for Storage in Deep Geologic Formations*, 1, 99-116, 2005

Eldrup, N. et al. (2019) 'A Cost Estimation Tool for CO2 Capture Technologies', SSRN Scholarly paper ID 3366036, doi: 10.2139/ssrn.3366036

Hasan, S. et al. (2021) 'Improving the Carbon Capture Efficiency for Gas Power Plants through Amine-Based Absorbents', *Sustainability*, 13(1), doi: 10.3390/su13010072

Kallevik, O. B. (2010) *Cost estimation of CO_2 removal in HYSYS*. Master's Thesis, Telemark University College, Porsgrunn.

Luo, X and Wang, M. (2016) 'Optimal operation of MEA-based post-combustion carbon capture for natural gas combined cycle power plants under different market conditions," *International Journal of Greenhouse Gas Control*, 48, pp. 312-320. doi: 10.1016/j.ijggc.2015.11.014.

Manzolini, G. et al. (2015) 'Economic assessment of novel amine based CO_2 capture technologies integrated in power plants based on European Benchmarking Task Force methodology'. *Applied Energy*, 138, pp. 546-548.

Mores, P. *et al.* (2012) 'CO_2 capture in power plants: Minimization of the investment and operating cost of the post-combustion process using MEA aqueous solution', *International Journal of Greenhouse Gas Control*, 10, pp. 148-163.

Nwaoha, C. *et al.* (2018), 'Techno-economic analysis of CO2 capture from a 1.2 million MTPA cement plant using AMP-PZ-MEA blend', *International Journal of Greenhouse Gas Control*, 78 pp. 400-412.

Park, K. and Øi, L. E. (2017) 'Optimization of gas velocity and pressure drop in CO_2 absorption column', Linköping Electronic Conference Proceedings SIMS 58, pp. 292-297. doi: 10.3384/ecp17138292

Roussanaly, S. (2019) "Calculating CO_2 avoidance costs of Carbon Capture and Storage from industry," *Carbon Management.*, 10(1), pp. 105–112. doi: 10.1080/17583004.2018.1553435.

Rubin, E. S. *et al*. (2013) 'A proposed methodology for CO2 capture and storage cost estimates', *International Journal of Greenhouse Gas Control*, 17, pp. 488-503.
Shirdel, S. *et al*. 'Sensitivity Analysis and Cost Estimation of a CO_2 Capture Plant in Aspen HYSYS'. *Chem. Engineering*, 6(2), 28, 2022.
Shirdel, S. (2022) *Process Simulation, Dimensioning and Automated Cost Optimization of CO_2 capture*. Master's Thesis, University of South-Eastern Norway, Porsgrunn.
Sulzer Chemtech, Separation Technology for the Hydrocarbon Processing Industry." Accessed: Oct. 28, 2021. [Online]. Available: https://www.sulzer.com/-/media/files/applications/oil-gas-chemicals/downstream/brochures/separation_technology_for_hpi.ashx
Van der Spek, M. *et al*. (2019) 'Best practices and recent advances in CCS cost engineering and economic analysis', *International Journal of Greenhouse Gas Control*, 83, pp. 91-104.
Øi, L. E. (2007) 'Aspen HYSYS simulation of CO_2 removal by amine absorption from a gas based power plant. In The 48th Scandinavian Conference on Simulation and Modeling (SIMS 2007); 30-31 October; 2007; *Linköping University Electronic Press*, 73-81, 2007.
Øi, L. E. (2012) *Removal of CO_2 from exhaust gas*. PhD Thesis, Telemark University College, Porsgrunn.
Øi, L. E. *et al*. (2021) 'Automated Cost Optimization of CO_2 Capture Using Aspen HYSYS', Linköping Electronic Conference Proceedings SIMS 62, pp. 293-300. doi: 10.3384/ecp21185293
Øi, L. E. *et al*. (2022) 'Process simulation and automated Cost Optimization of CO_2 Capture Using Aspen HYSYS', Linköping Electronic Conference Proceedings SIMS 63, pp. 65-71.

The Effect of Climate and Orientation on the Energy Performance of a Prefab House in Norway

Amirhossein Ghazi *, Zahir Barahmand, Lars Erik Øi

Department of Process, Energy and Environmental Technology, University of South-Eastern Norway
259023@usn.no

Abstract

Norway has a wide range of climatic conditions throughout the country. The climate varies from coastal to inland areas. Geographic latitude and longitude, as well as the gulf stream oceanic flow, account for this phenomenon. Different climate types can certainly affect residential building heating energy demands and make overheating more likely. On the other hand, a building's orientation has an impact on its heating energy requirements. A building's orientation affects how much solar gain it receives and how much wind it receives over the course of the year. Employing DesignBuilder® software, This study examines how different orientations affect the energy performance of a pre-designed house with and without solar photovoltaic panels in typical Norwegian climates. The results confirm that in different locations, the optimal situation is South-East and the lowest energy consumption without and with photovoltaic panels belongs to Bergen with 83305 Wh/m^2 and Oslo with 29442 Wh/m^2 respectively. This comparative study will be helpful to stakeholders in the building ecosystem (municipalities, engineers, and designers, building companies, suppliers, and residents) in making more informed decisions.

1. Introduction

Norwegian households used 48 TWh of energy, or 22 percent of the total energy consumed (Statistics Norway, 2021). The non-residential sector contributes almost 18% of total energy which implies that almost 40 percent of the final energy consumption in Norway comes from the building stock (Sartori et al., 2009). Generally, these statistics apply to western countries as well (*EU Energy and Transport in Figures*, 2009). Efforts are being made by the Norwegian authorities to reduce the energy demand for buildings (Korsnes et al., 2013). Recent revisions to the technical building regulations (TEK17, 2017) require greater insulation, heat recovery, and airtightness than earlier versions.

In the household as well as in the service sector, electricity is the most widely used energy carrier (Fig. 1). Electricity has been increasing in the energy mix, reaching 83% in 2017 and this confirms the importance of the possibility of generating electricity from the house itself via solar panels. The second largest portion of household energy is derived from biofuels. About 5.8 TWh of energy was generated by biofuels in 2017. Fuelwood constitutes the majority of this energy, but pellets and bio-oils are also used by households. (“Energy use by sector”)

Often referred to as prefab or modular homes, prefabricated houses are manufactured off-site and then transported to the building site for final assembly. Assembling prefabricated houses involves precutting and prefabricating building components, such as walls, roofs, floors, and doors, in a controlled environment before they are transported to the building site. In general, predesigned and prefabricated homes offer several advantages, including lower costs, energy efficiency, and versatility, making them an increasingly popular choice in many countries. Due to these factors, Scandinavia and Norway have a long history of using prefabricated houses.

There can be a significant impact on the amount of energy required to heat and cool a house based on the local climate. In cooler climates, for instance, homes require more energy to stay warm in the winter, while in warmer climates, homes require more energy to stay cool in the summer. On the other hand, a house's orientation can also affect its energy efficiency. It is possible to reduce the amount of energy required for heating a residence with large windows facing the south during the winter months by utilizing natural solar heat gain. South-facing windows, however, can increase heat absorption by the home in hotter climates, which increases cooling energy requirements. In addition to influencing the amount of natural light that enters a home, the orientation can also have a significant impact on the energy consumption of the home by reducing the need for artificial light.

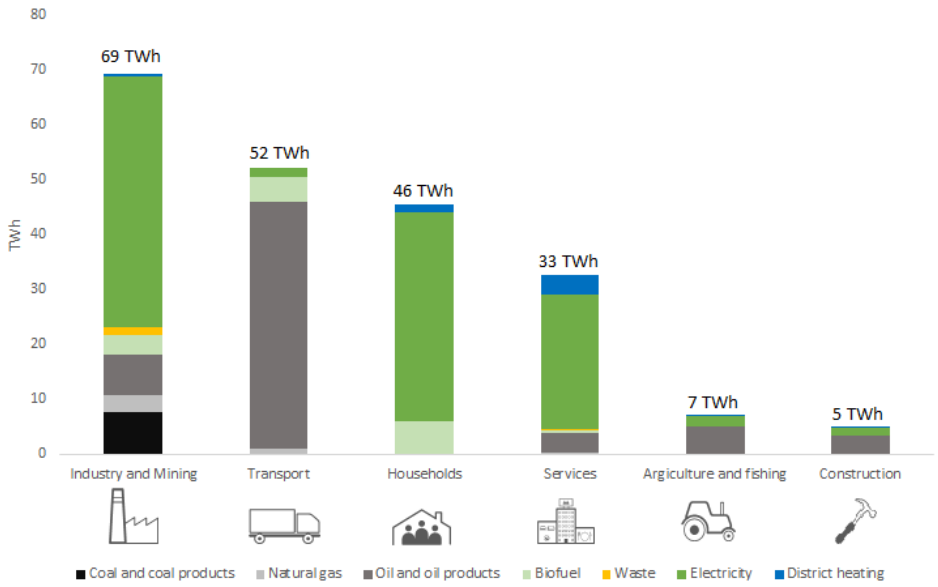


Figure 1: Norway's final energy consumption by energy carrier in 2020 (adapted from ("Energy use by sector"))

In the present study, a pre-designed house in five different typical Norwegian climates was tested to answer the following: 1) In different climates, how does a house's orientation affect its energy efficiency? 2) What is the optimal orientation for the house in the selected location? 3) What is the effect of climate on energy consumption? 4) What is the effect of solar photovoltaic panels on the energy consumption of a house in different climates? A bioclimatic paradigm was used to assess the operational energy and daylight performance of this type of building in various climates.

2. Literature Review

Nordic countries' cold climate and abundance of natural resources have created unique challenges and opportunities in the field of energy efficiency. This has resulted in a significant amount of research being conducted on the energy performance of these countries (Abrahamsen et al., 2023; Carpino et al., 2020, 2020; Cohen et al., 2007; Liu et al., 2015; Mahapatra & Olsson, 2015; Molin et al., 2011; Tommerup et al., 2007).

Many factors can affect the energy efficiency of a house, including the climate (Cronin et al., 2018; Li et al., 2021) and its orientation (Abanda & Byers, 2016; Albatayneh et al., 2018; Elghamry & Azmy, 2017; Lahmar et al., 2022). There have been numerous studies that examine the effects of climate, building orientation, location, etc. on a building's energy performance. In particular, the orientation of the façade has a significant impact on the performance of building integrated photo voltaic (BIPV) on façades (Akbarinejad et al., 2022). As a consequence of the relatively symmetrical sun path throughout the day, it is difficult to determine the of a building located in a warm-humid climate. Nicoletti et al. (2022) employed EnergyPlus to evaluate the energy and visual performance of a building with photochromic glazings in southern Italy. By considering five climatic locations in Saudi Arabia, Alyami et al. (2022) examined the effects of location and insulation material on the energy efficiency of residential buildings. By observing and conducting experiments on four existing buildings, Khaliq and Mansoor (2022) determined the effectiveness of energy consumption, as well as developing a model based on different contributing parameters, including orientation, construction materials, construction type, etc. Morsali et al. (2021) investigated the effects of building direction and roofs on the energy consumption of residential buildings through simulations using Building Information Models. Abdul Mujeebu and Ashraf (2020) examined the impact of location and range of thermostat set points for cooling and heating on nano gel glazing energy performance and economics in a multistory office building, considering 26 climatic regions across Saudi Arabia. Various climate regions, Lapsia (2019) investigated the effect of its geometric shape and orientation on its energy performance. Fela et al. (2019) evaluated the impact of climate on daylight performance in a reference office in which there is only one glazed opening, and on which a range of window-to-wall ratios are measured on one of the short façades facing a variety of orientations. Tab. 1 summarizes the results from the literature review. optimal façade orientation for tropical cities in terms of maximum energy yield and daylight performance. On tropical building façades, Mangkuto et al. (2023) determined the optimum orientation for BIPV. Karthick et al. (2023) examined the effects of building orientation, window glazing, and shading techniques on the energy efficiency and comfort.

Table 1: Summary of literature review

References	Main Parameter	Climate	Software	Location
(Mangkuto et al., 2023)	Building Orientation	Tropical		Indonasia
(Karthick et al., 2023)	Building Orientation, Window Glazing, and Shading Techniques	Warm-Humid	DesignBuilder, EnergyPlus	
(Nicoletti et al., 2022)	Building Location, Window Glazing		EnergyPlus	Southern Italy
(Alyami et al., 2022)	Building Location and Insulation Material	Hot-Humid	IES-VE	Saudi Arabia
(Khaliq & Mansoor, 2022)	Orientation, construction materials, type of construction			Pakistan
(Morsali et al., 2021)	Roof shapes and building orientation		REVIT	
(Abdul Mujeebu & Ashraf, 2020)	Location and deadband	Hot-Humid	Ecotect	Saudi Arabia
(Baruah & Sahoo, 2020)	Orientations, location types of roof surfaces, walls and fenestrations	Sub-tropical humid climate with dry winter conditions	eQUEST	Himalayan terrain of India
(Lapisa, 2019)	Different climates, geometric shapes, and orientation		TRNSYS, CONTAM	Jakarta, Marseille, and Poitiers
(Hammad et al., 2018)	Location and design of windows		Green Building Studio	The middle east and north africa (MENA)
(Dobosi et al., 2019)	Various locations		EnergyPlus in Sketchup	Romania
(Fela et al., 2019)	Orientation, window size, and lighting control	Tropical area	Radiance and Daysim	Indonesia
(Košir et al., 2018)	Location's climatic specifics		EnergyPlus in OpenStudio	
(Elhadad et al., 2018)	Building orientation		IDA ICE 4.7	Egypt
(Khan & Asif, 2017)	Green roof and building orientation	Hot-Humid	Ecotect	Saudi Arabia
(Poddar et al., 2017)	Building orientations and seasonal variations		DesignBuilder, EnergyPlus	South Korea
(Diaz & Osmond, 2017)	Various locations	Hot-Humid Tropics	WUFI Plus	Location

3. Simulation Setup

A pre-designed two-story Norwegian house (Fig. 2) was designed in accordance with the spaces and areas specified in Tab. 2.

To meet the requirements of the Norwegian regulations, the materials used in this house have been selected so as to meet the requirements of the TEK 17 standard (*Byggteknisk Forskrift (TEK17) Med Veiledning*, 2017), which has been adopted by the Norwegian government. A summary of the requirements for external walls, roofs, floors, and windows prescribed by TEK 17 can be found in Tab. 3.

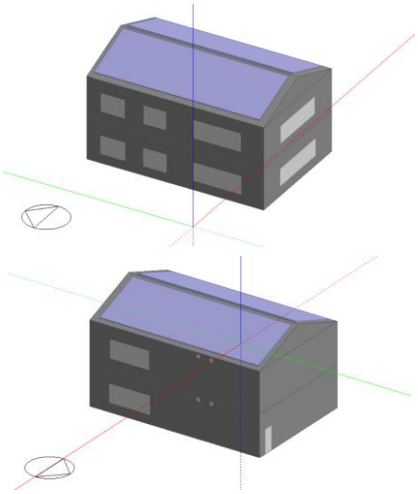


Figure 2: The overview of the building

Table 2: Names and areas of the building spaces

	Name of Areas	**Area (m^2)**
First story	Living Room	85.825
	Kitchen	10.954
	Toilet-Bathroom	5.995
	Bedroom 1	20.289
	Bedroom 2	17.351
	Hallway	10.144
Second story	Living Room	85.825
	Kitchen	10.954
	Toilet-Bathroom	5.995
	Bedroom 1	20.289
	Bedroom 2	17.351
	Hallway	10.144
	Total Area of Each Story	**150.558**
	Total Area of the Building	**301.116**

Table 3: Comparison of standard requirements and used values for the simulations

Small house (150 m^2)	**Requirement of TEK 17**	**Used**
U-value outer walls [W/(m^2 K)]	≤ 0.18	0.176
U-value roof [W/(m^2 K)]	≤ 0.13	0.127
U-value floors [W/(m^2 K)]	≤ 0.10	0.094
U-value windows and doors [W/(m^2 K)]	≤ 0.80	0.78
Proportion of window and door areas of heated gross internal area	≤ 25%	≤ 25%

The heating system of the house is ground heating fed by a hot water boiler which uses electricity from the grid to heat the water and has a coefficient of performance (CoP) equal to 0.65 (the default CoP specified in the software's library for this system). Due to the climate characteristics, no cooling system is considered for the house.

Noteworthy to indicate is that the ventilation rate has been set to 0.5 air change per hour (the minimum permitted amount) (Dimitroulopoulou, 2012) and the air infiltration rate has been set to 0.3 air change per hour as it must be under 0.6 (Bunkholt et al., 2021).

Moreover, to check the possibility and the amount of electricity generation by solar energy in each city, the pitched roof of the building is covered with solar photovoltaic (PV) panels with characteristics such as area of the PV panel equal to 128 m^2, efficiency of 0.15, and fraction of surface with active solar cells equal to 0.9.

Similar to many other studies and simulations, our study has also limitations. There are other parameters to be set based on the DesignBuilder® software requirements which have been set as the default value of the software itself. Moreover, as mentioned above, there are some critical parameters such as CoP, air infiltration and ventilation rates which the results are sensitive to them, so they are worthy to be studied in the future.

From another point of view, as seen in the next section, only eight main orientations have been considered in the simulations. And necessarily the optimum orientation is not among these. In addition, the shading effect of other buildings has not been considered in this study and the slope of the PV cells has been set as the slope of the roof which is not necessarily optimum.

4. Results and Discussion

As a result of setting up all the above parameters in the DesighnBuilder® software, five Norwegian cities with differing climates were selected as the locations for the house, namely Oslo, Trondheim, Tromsø, Kristiansand, and Bergen. Throughout each city, one simulation has been performed for each direction (south, south-east, east, north-east, north, north-west, west, south-west). To calculate the energy consumption and energy generation of PV panels during the year, as well as the percentage of energy consumption reduction with generation during the year, the house was rotated in the eight directions listed above. As a result, the data referred to above were exported and are shown in Appendix A.

The simulations show that the lowest energy consumption in each city can be reached in the facing into the direction of south-east with the amount of 87715, 102530, 135250, 87759, 83305

[Wh/m^2] per year for Oslo, Trondheim, Tromsø, Kristiansand, and Bergen respectively (Fig. 3).

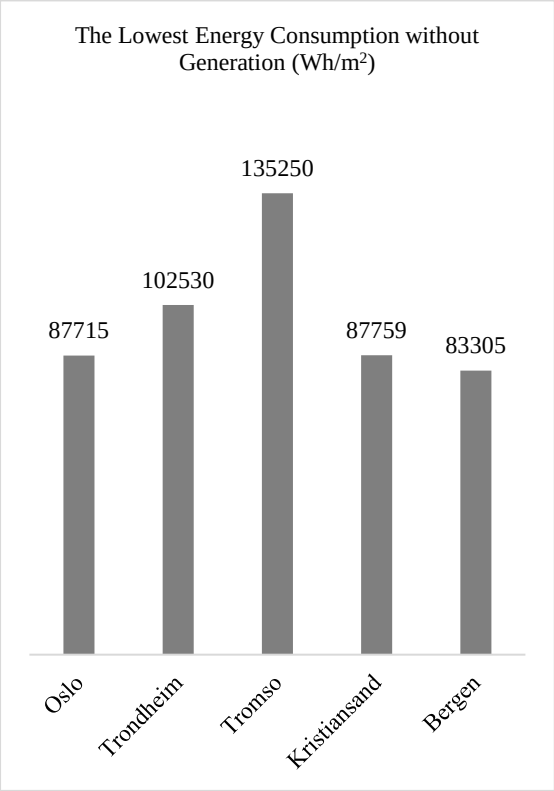


Figure 3: The lowest energy consumption in the cities among all orientations

Furthermore and as it could also be expected, Tromsø and Trondheim have the highest yearly energy consumption, which is the result of their geographical location and climate.

Interestingly, the highest energy generation with the PV panels is available in a different facing in comparison to the facing with the lowest energy consumption in each city. The PV panels are able to produce the highest amount of energy in the facing into the West with the amount of 58366, 58920, 49770, 58366, and 49073 [Wh/m^2] per year for Oslo, Trondheim, Tromsø, Kristiansand, and Bergen respectively (Fig. 4). It is expected that the performance of the PV panels are highly affected negatively in cloudy weathers such as Bergen and Tromsø.

Although it is not surprising that the second lowest amount of energy generation by PV panels is in Tromsø because of the angle of the sun light due to the altitude of the city and because of the long periods of darkness, it is absolutely surprising that Bergen has the first lowest amount of energy generation while it has considerably lower altitude and also has shorter periods of darkness comparing to Tromsø.

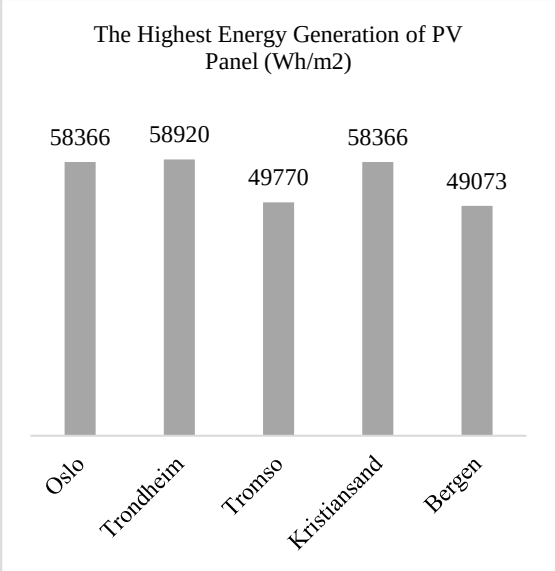


Figure 4: The highest energy generation with the PV panels in the cities among all orientations

By combining the yearly energy consumption with and without energy generation by PV panels, it can be concluded that by adding PV panels to this building, the maximum energy reduction can be achieved in the south-east facing by 66.43, 57.42, 36.7, 66.4, and 58.85 percent for Oslo, Trondheim, Tromsø, Kristiansand, and Bergen respectively (Fig. 5).

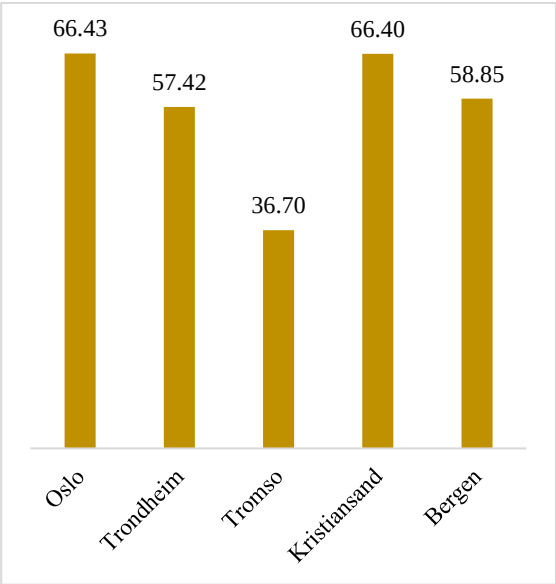


Figure 5: The highest energy consumption reduction using PV panel (%) in each city among all arientations

Finally, although the highest energy generation is achieved in Trondheim (58920 Wh/m^2), the highest percentage of energy consumption reduction with generation is achieved in Oslo and Kristiansand with a negligible difference, 66.43 and 66.4 percent respectively (Fig. 6).

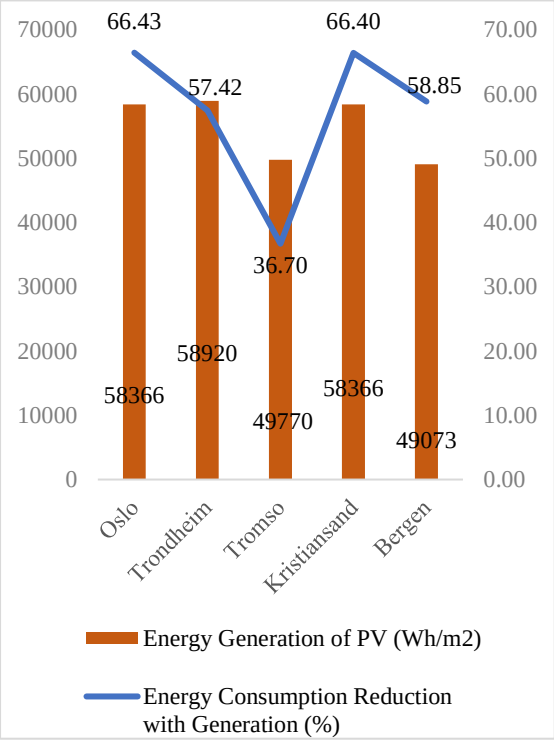


Figure 6: The highest energy generation and energy consumption reduction in the cities among all orientations

5. Conclusion

The findings of the study highlighted the significance of considering climate variations and building orientation when assessing residential building heating energy demands and the likelihood of overheating. The geographic latitude and longitude, along with the influence of the Gulf Stream oceanic flow, were identified as contributing factors to the diverse climate types across Norway.

Through the analysis, it was determined that the optimal orientation for energy efficiency differed across locations, with the South-East direction emerging as the most favorable in the examined scenarios. The study also investigated the impact of incorporating solar photovoltaic panels into the house design, noting that different locations demonstrated varying energy consumption levels. The city of Bergen showed the lowest energy consumption without photovoltaic panels, recording 83,305 Wh/m^2, while Oslo exhibited the lowest consumption with photovoltaic panels, at 29,442 Wh/m^2.

By shedding light on these energy performance variations, this study provides valuable guidance for stakeholders involved in the design and construction of residential buildings in Norway.

Appendix A: The extracted data from simulations

City	Facing	Energy consumption without generation (Wh/m^2)	Energy consumption difference according to minimum amount (%)	Energy generation of PV (Wh/m^2)	Energy consumption with generation (Wh/m^2)	Energy consumption reduction with generation (%)
Oslo	S	89343	1.86	58221	31122	65.17
	SE	87715	0.00	58273	29442	66.43
	E	88242	0.60	58333	29909	66.11
	NE	92290	5.22	58225	34065	63.09
	N	95970	9.41	58222	37748	60.67
	NW	95260	8.60	58297	36963	61.20
	W	92760	5.75	58366	34394	62.92
	SW	90810	3.53	58248	32562	64.14
Trondheim	S	104210	1.64	58890	45320	56.51
	SE	102530	0.00	58870	43660	57.42
	E	103260	0.71	58880	44380	57.02
	NE	107650	4.99	58850	48800	54.67
	N	111690	8.93	58890	52800	52.73
	NW	111410	8.66	58900	52510	52.87
	W	108570	5.89	58920	49650	54.27
	SW	106100	3.48	58880	47220	55.49

Tromsø	S	136980	1.28	49540	87440	36.17
	SE	135250	0.00	49640	85610	36.70
	E	135990	0.55	49740	86250	36.58
	NE	140730	4.05	49600	91130	35.24
	N	144640	6.94	49540	95100	34.25
	NW	144050	6.51	49670	94380	34.48
	W	141160	4.37	49770	91390	35.26
	SW	138890	2.69	49620	89270	35.73
Kristiansand	S	89395	1.86	58221	31174	65.13
	SE	87759	0.00	58273	29486	66.40
	E	88285	0.60	58333	29952	66.07
	NE	92370	5.25	58225	34145	63.03
	N	96030	9.42	58222	37808	60.63
	NW	95320	8.62	58297	37023	61.16
	W	92780	5.72	58366	34414	62.91
	SW	90860	3.53	58248	32612	64.11
Bergen	S	84359	1.27	49040	35319	58.13
	SE	83305	0.00	49023	34282	58.85
	E	83555	0.30	49051	34504	58.71
	NE	86045	3.29	49034	37011	56.99
	N	88497	6.23	49040	39457	55.41
	NW	88300	6.00	49039	39261	55.54
	W	86681	4.05	49073	37608	56.61
	SW	85362	2.47	49049	36313	57.46

References

Abanda, F. H., & Byers, L. (2016). An investigation of the impact of building orientation on energy consumption in a domestic building using emerging BIM (Building Information Modelling). *Energy*, *97*, 517–527. https://doi.org/10.1016/j.energy.2015.12.135

Abdul Mujeebu, M., & Ashraf, N. (2020). Impact of location and deadband on energy performance of nano aerogel glazing for office building in Saudi Arabia. *Building Research & Information*, *48*(6), 645–658. https://doi.org/10.1080/09613218.2019.1696171

Abrahamsen, F. E., Ruud, S. G., & Gebremedhin, A. (2023). Assessing Efficiency and Environmental Performance of a Nearly Zero-Energy University Building's Energy System in Norway. *Buildings*, *13*(1). Scopus. https://doi.org/10.3390/buildings13010169

Akbarinejad, T., Barahmand, Z., & Samarakoon, G. (2022). *Simulation-Based Life Cycle Assessment for Office Building Façade: A Case Study of the Leadenhall Building in London*. https://doi.org/10.3384/ecp192019

Albatayneh, A., Alterman, D., Page, A., & Moghtaderi, B. (2018). The Significance of the Orientation on the Overall buildings Thermal Performance-Case Study in Australia. *Energy Procedia*, *152*, 372–377. https://doi.org/10.1016/j.egypro.2018.09.159

Alyami, S. H., Alqahtany, A., Ashraf, N., Osman, A., Aldossary, N. A., Almutlaqa, A., Al-Maziad, F., Alshammari, M. S., & Al-Gehlani, W. A. G. (2022). Impact of Location and Insulation Material on Energy Performance of Residential Buildings as per Saudi Building Code (SBC) 601/602 in Saudi Arabia. *Materials*, *15*(24), Article 24. https://doi.org/10.3390/ma15249079

Baruah, A., & Sahoo, S. (2020). Energy efficiency performance analysis of a residential building for the effects of building orientations, types of roof surfaces, walls and fenestrations at different locations in the Himalayan terrain of India. *AIP Conference Proceedings*, *2273*(1), 050061. https://doi.org/10.1063/5.0024245

Bunkholt, N. S., Gullbrekken, L., Time, B., & Kvande, T. (2021). Process induced building defects in Norway – development and climate risks. *Journal of Physics: Conference Series*, *2069*(1), 012040. https://doi.org/10.1088/1742-6596/2069/1/012040

Byggteknisk forskrift (TEK17) med veiledning. (2017). Direktoratet for Byggkvalitet. https://dibk.no/regelverk/byggteknisk-forskrift-tek17

Carpino, C., Loukou, E., Heiselberg, P., & Arcuri, N. (2020). Energy performance gap of a nearly Zero Energy Building (nZEB) in Denmark: The influence of occupancy modelling. *Building Research and Information*, *48*(8), 899–921. Scopus. https://doi.org/10.1080/09613218.2019.1707639

Cohen, R., Bordass, W., & Leaman, A. (2007). *Evaluations and comparisons of the achieved energy and environmental performance of two library buildings in England and Sweden*. *113 PART 2*, 14–26. Scopus.

Cronin, J., Anandarajah, G., & Dessens, O. (2018). Climate change impacts on the energy system: A review of trends and gaps. *Climatic Change*, *151*(2), 79–93. https://doi.org/10.1007/s10584-018-2265-4

Diaz, C. A., & Osmond, P. (2017). Influence of Rainfall on the Thermal and Energy Performance of a Low Rise Building in Diverse Locations of the Hot Humid Tropics. *Procedia Engineering*, *180*, 393–402. https://doi.org/10.1016/j.proeng.2017.04.198

Dimitroulopoulou, C. (2012). Ventilation in European dwellings: A review. *Building and Environment*, *47*, 109–125. https://doi.org/10.1016/j.buildenv.2011.07.016

Dobosi, I. S., Tanasa, C., Kaba, N.-E., Retezan, A., & Mihaila, D. (2019). Building energy modelling for the energy performance analysis of a hospital building in various locations. *E3S Web of Conferences*, *111*, 06073. https://doi.org/10.1051/e3sconf/201911106073

Elghamry, R., & Azmy, N. (2017). *Buildings orientation and it's impact on the energy consumption*.

Elhadad, S., Baranyai, B., & Gyergyák, J. (2018). The impact of building orientation on energy performance: A case study in New Minia, Egypt. *Pollack Periodica*, *13*, 31–40. https://doi.org/10.1556/606.2018.13.3.4

Energy use by sector. (n.d.). Energifakta Norge. Retrieved 19 March 2023, from https://energifaktanorge.no/en/norsk-energibruk/energibruken-i-ulike-sektorer/

EU energy and transport in figures. (2009). Office for Official Publications of the European Communities.

Fela, R., Utami, S., Mangkuto, R., & Suroso, D. (2019). *The Effects of Orientation, Window Size, and Lighting Control to Climate-Based Daylight Performance and Lighting Energy Demand on Buildings in Tropical Area*. https://doi.org/10.26868/25222708.2019.210677

Hammad, A., Akbarnezhad, A., Grzybowska, H., Wu, P., & Wang, X. (2018). Mathematical optimisation of location and design of windows by considering energy performance, lighting and privacy of buildings. *Smart and Sustainable Built Environment*, *8*(2), 117–137. https://doi.org/10.1108/SASBE-11-2017-0070

Karthick, S., Sivakumar, A., Bhanu Chander, S., Bharathwaj, S., & Kumar, B. V. V. N. (2023). Effect of Building Orientation, Window Glazing, and Shading Techniques on Energy Performance and Occupant Comfort for a Building in Warm-Humid Climate. In N. L. Narasimhan, M. Bourouis, & V. Raghavan (Eds.), *Recent Advances in Energy Technologies* (pp. 289–310). Springer Nature. https://doi.org/10.1007/978-981-19-3467-4_18

Khaliq, W., & Mansoor, U. B. (2022). Performance evaluation for energy efficiency attainment in buildings based on orientation, temperature, and humidity parameters. *Intelligent Buildings International*, *14*(5), 606–622. https://doi.org/10.1080/17508975.2021.1873096

Khan, H. S., & Asif, M. (2017). Impact of Green Roof and Orientation on the Energy Performance of Buildings: A Case Study from Saudi Arabia. *Sustainability*, *9*(4), Article 4. https://doi.org/10.3390/su9040640

Korsnes, S., Time, B., Simonsen, I., & Kristensen, K. (2013). The climatic challenge of designing a prefabricated catalogue house. *Proceeding of Passivhus Norden*. the 6th Passive House Conference in the Nordic Countries.

Košir, M., Iglič, N., & Kunič, R. (2018). Optimisation of heating, cooling and lighting energy performance of modular buildings in respect to location's climatic specifics. *Renewable Energy*, *129*, 527–539. https://doi.org/10.1016/j.renene.2018.06.026

Lahmar, I., Cannavale, A., Martellotta, F., & Zemmouri, N. (2022). The Impact of Building Orientation and Window-to-Wall Ratio on the Performance of Electrochromic Glazing in Hot Arid Climates: A Parametric Assessment. *Buildings*, *12*(6), Article 6. https://doi.org/10.3390/buildings12060724

Lapisa, R. (2019). The effect of building geometric shape and orientation on its energy performance in various climate regions. *International Journal of GEOMATE*, *16*. https://doi.org/10.21660/2019.53.94984

Li, Y., Wang, W., Wang, Y., Xin, Y., He, T., & Zhao, G. (2021). A Review of Studies Involving the Effects of Climate Change on the Energy Consumption for Building Heating and Cooling. *International Journal of Environmental Research and Public Health*, *18*(1), 40. https://doi.org/10.3390/ijerph18010040

Liu, M., Wittchen, K. B., & Heiselberg, P. K. (2015). Control strategies for intelligent glazed façade and their influence on energy and comfort performance of office buildings in Denmark. *Applied Energy*, *145*, 43–51. Scopus. https://doi.org/10.1016/j.apenergy.2015.02.003

Mahapatra, K., & Olsson, S. (2015). Energy performance of two multi-story wood-frame passive houses in Sweden. *Buildings*, *5*(4), 1207–1220. Scopus. https://doi.org/10.3390/buildings5041207

Mangkuto, R. A., Tresna, D. N. A. T., Hermawan, I. M., Pradipta, J., Jamala, N., Paramita, B., & Atthaillah. (2023). Experiment and simulation to determine the optimum orientation of building-integrated photovoltaic on tropical building façades considering annual daylight performance and energy yield. *Energy and Built Environment*. https://doi.org/10.1016/j.enbenv.2023.01.002

Molin, A., Rohdin, P., & Moshfegh, B. (2011). Investigation of energy performance of newly built low-energy buildings in Sweden. *Energy and Buildings*, *43*(10), 2822–2831. https://doi.org/10.1016/j.enbuild.2011.06.041

Morsali, S., Akbarian, S., & Hamed Zar Gari, Z. (2021). Simulation of the roof shapes and building orientation on the energy performance of the buildings. *Journal of Building Pathology and Rehabilitation*, *6*(1), 36. https://doi.org/10.1007/s41024-021-00132-3

Nicoletti, F., Kaliakatsos, D., Ferraro, V., & Cucumo, M. A. (2022). Analysis of the energy and visual performance of a building with photochromic windows for a location in southern Italy. *Building and Environment*, *224*, 109570. https://doi.org/10.1016/j.buildenv.2022.109570

Poddar, S., Park, D., & Chang, S. (2017). Energy performance analysis of a dormitory building based on different orientations and seasonal variations of leaf area index. *Energy Efficiency*, *10*(4), 887–903. https://doi.org/10.1007/s12053-016-9487-y

Sartori, I., Wachenfeldt, B. J., & Hestnes, A. G. (2009). Energy demand in the Norwegian building stock: Scenarios on potential reduction. *Energy Policy*, *37*(5), 1614–1627. https://doi.org/10.1016/j.enpol.2008.12.031

Statistics Norway. (2021). *Production and consumption of energy, energy balance and energy account*. SSB. https://www.ssb.no/en/energi-og-industri/energi/statistikk/produksjon-og-forbruk-av-energi-energibalanse-og-energiregnskap

TEK17. (2017). *Building technical regulations* (Regulation FOR-2017-06-19-840). Ministry of Local Government and Districts.

Tommerup, H., Rose, J., & Svendsen, S. (2007). Energy-efficient houses built according to the energy performance requirements introduced in Denmark in 2006. *Energy and Buildings*, *39*(10), 1123–1130. https://doi.org/10.1016/j.enbuild.2006.12.011

Approaching simulation-based controller design: heat exchanger case study

Matias Waller[a,b,*] Leonardo Espinosa-Leal[b]

[a] *Åland University of Applied Sciences, Mariehamn, Åland;* [b] *Arcada University of Applied Sciences, Helsinki, Finland*

[*] matias.waller@arcada.fi

Abstract

Supported by an identification experiment using random-phase multisines, uncertain parameters in a grey-box model for a multiple-input multiple-output laboratory-scale heat exchanger are fitted to experimental data. By defining desired trajectories for the controlled system concerning setpoint changes, simulations and a cost function taking control signal activity into account, we determine both a linear and a nonlinear PI-controller. The resulting control systems are evaluated through practical experiments and analysis with encouraging results. The approach to modelling and controller design raises questions about what is needed from an educational point of view, e.g., what skills are needed for simulation-based control design and analysis?

1 Introduction

The development of increasingly accurate models and simulations raises questions about whether new approaches to the design of control systems should be considered. As an example, models often popularly referred to as Digital Twins (DTs) (Espinosa-Leal et al., 2020) that combine sensor information, data-based black-box models with physical models to develop a faithful virtual replica of a given system are used onboard in the electronic control unit for improving the final tuning phase of the controller for the physical vehicle in an automobile application (Dettù et al., 2023). Inspired by similar approaches and the approach of (nominal) model predictive control (mpc) in general (Hewing et al., 2020), we explore possibilities to use simulations directly for controller design. Our case study considers a laboratory-scale heater and a heat exchanger. Heat exchangers are vital components in, among other applications, thermal power stations and heat pumps. Improved control of heat exchangers based on the creative use of increasingly accurate models could thus contribute to the much needed increased energy efficiency (IPCC, 2023).

Although it would be appealing to illustrate the approach on a deterministic, linear model that could be fully explored analytically, the primary motivation for using simulations directly is, naturally, for processes with features not easily captured by linearized models. Therefore, the approach is illustrated in a nonlinear, multivariable process. In this initial approach, we, for manageability, focus on the control of one vital quality variable. Combining first-principles with experiments, a system of coupled nonlinear differential equations forms the basic model of the heating process considered. Some uncertain parameters of the model are fitted based on an identification experiment. A sequence of setpoint changes and a control performance criterion are then defined. Different control strategies are explored and optimized based on simulations. The strategies are then implemented in practice, results are evaluated, and a preliminary analysis of the control systems is presented.

Reflecting on the approach, what does it imply regarding modelling and model analysis required for teaching control engineering on a general level? If control design is based on simulations only, does this mean that the skills needed for analyzing (and linearizing) differential or difference equations are less critical? How can traditional requirements on stability analysis, robustness, control performance and control signal activity be explored based on simulations? Connected to these questions is the opaque nature of general nonlinear black-box models and physical component models with block diagrams that often make them less suitable for traditional, linear approaches. The gaining popularity of these modelling alternatives also motivates the approach presented in this paper.

2 Laboratory-scale heat exchanger

A schematic view of the process is illustrated in the left panel of Fig. 1. The process has three control signals, u_c controlling the pump on the cold side, u_h controlling the pump on the hot side, and u_p controlling the power to the heater. Given the external disturbances

T_{ci} and T_{surr}, i.e., the temperature of the flow on the cold side into the heat exchanger and the surrounding temperature, a model should be able to determine the power to the heater P, the flows on the cold and on the hot side, $\dot{V}_c$ and $\dot{V}_h$, the temperatures entering and exiting on the hot side, T_{hi} and T_{ho}, and the temperature exiting on the cold side, T_{co}.

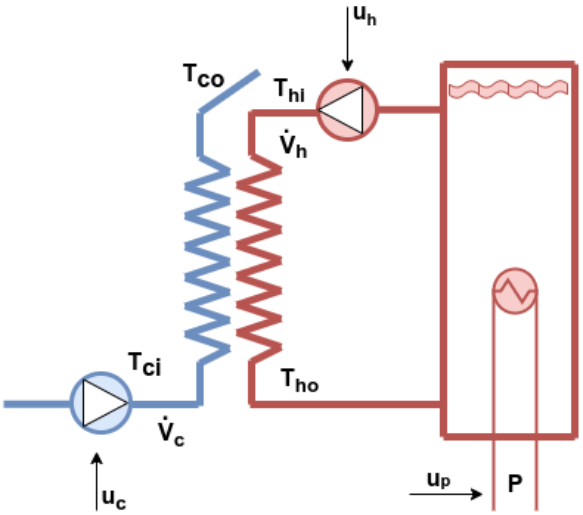


Figure 1. Schematic view of the heating process. Blue denotes the cold side of the heat exchanger, and red denotes the hot side and the heater.

The main objective is to control $\dot{V}_c$ as well as T_{co}. The temperatures on the hot side are less interesting, but naturally sufficient heat is necessary to enable control of T_{co}.

2.1 A grey-box model

In the approach to the simulation-based design of controllers, we consider a simplified first-principles model for the heating process. A net energy balance can be written

$$\frac{dE_{tot}}{dt} = P - \dot{V}_c \rho c_p (T_{co} - T_{ci}) - \dot{Q}_{loss} \qquad (1)$$

where dE_{tot}/dt is the change of net stored energy in the system and $\dot{Q}_{loss}$ are the heat losses to the environment. The term E_{tot} cannot easily be used to determine the temperatures T_{hi}, T_{ho} and T_{co}. For this reason, the net balance is split into three equations. A balance over the heater yields

$$\frac{dE_p}{dt} = P - \dot{V}_h \rho c_p (T_{hi} - T_{ho}) - \dot{Q}_{loss} \qquad (2)$$

where all heat losses from the system are assigned to the heater. The stored energy is assumed to be characterized by the temperature T_{hi} and given by $E_p = C_p(T_{hi} - T_{ref})$ where C_p is the heat capacity for the heater and T_{ref} is a reference temperature. If C_p is assumed constant, $dE_p/dt = C_p dT_{hi}/dt$. A similar balance over the hot side of the heat exchanger yields

$$\frac{dE_h}{dt} = \dot{V}_h \rho c_p (T_{hi} - T_{ho}) - \dot{Q}_{he} \qquad (3)$$

where $\dot{Q}_{he}$ is the power transferred in the heat exchanger and dE_h/dt is characterized by T_{ho}. If C_h is the heat capacity for the hot side of the heat exchanger and assumed constant, $dE_h/dt = C_h dT_{ho}/dt$. The corresponding balance over the cold side of the heat exchanger is given by

$$\frac{dE_c}{dt} = \dot{Q}_{he} - \dot{V}_c \rho c_p (T_{co} - T_{ci}) \qquad (4)$$

where dE_c/dt is characterized by the temperature T_{co}. If C_c is the heat capacity for the cold side of the heat exchanger and assumed constant, $dE_c/dt = C_c dT_{co}/dt$. In addition to these energy balances, equations for $\dot{Q}_{loss}$, $\dot{Q}_{he}$ and equations for dependencies between u_c and $\dot{V}_c$, u_h and $\dot{V}_h$ and u_p and P are needed.
For a heat exchanger, it is common to use $\dot{Q}_{he} = \alpha A \Delta T_{lm}$, where α is the heat transfer coefficient, A is the exchange area and ΔT_{lm} is the logarithmic mean temperature difference. In this case, the temperature differences are $T_{hi} - T_{co}$ and $T_{ho} - T_{ci}$. Although ΔT_{lm} is motivated by steady-state, we nonetheless use it for our dynamic model. The heat losses are assumed to be proportional to $T_{hi} - T_{surr}$, i.e., $\dot{Q}_{loss} = k(T_{hi} - T_{surr})$.
For the pumps and the heater, first-order models were fitted to step experiments to give the equations,

$$\begin{aligned} \frac{d\dot{V}_c}{dt} &= \frac{1}{T_c}(K_c \max(u_c - 19, 0) - \dot{V}_c) \\ \frac{d\dot{V}_h}{dt} &= \frac{1}{T_h}(K_h \max(u_h - 19, 0) - \dot{V}_h) \\ \frac{dP}{dt} &= \frac{1}{T_p}(K_p u_p - P) \end{aligned} \qquad (5)$$

with $T_c = 1.5$ s, $K_c = 5$ (ml/(min)/%), $T_h = 1$ s, $K_h = 10$ (ml/min)/%), $T_p = 1$ s and $K_p = 0.016$ kW/%.
The models include several uncertain parameters in addition to the pump and power characteristics given above, mainly α, k, C_c, C_h, and C_p. Constant estimates for these will, for simplicity, be considered. In addition, delays are determined by visual inspection of step changes and possible flow-dependent transport delays as well as distribution of temperatures and flows, ageing and other time-variant characteristics are neglected. The identification experiment and how the parameters α, k, C_c, C_h and C_p are fitted is described next.

2.2 Identification experiment

Although α and k could be fitted to steady-state data, estimating heat capacities requires experiments with

nonzero temperature derivatives. In future work, we also aim to explore various data-based models for simulation-based controller design. For this reason, a 6000 seconds identification experiment using random-phase multisines (Pintelon & Schoukens, 2012) for all three control signals, u_c, u_h and u_p, was conducted. The main experiment contains two independent sets of three 1000 seconds long periods each. A short segment of the experiment illustrating T_{co}, corresponding simulations $T_{\text{co},s}$ and the three control signals are illustrated in Fig. 2.

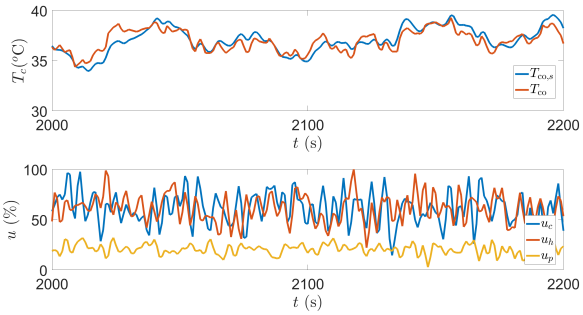


Figure 2. Upper panel: Simulations of $T_{\text{co},s}$ (blue) and measurements T_{co} (red) for a short segment of the identification experiment. Lower panel: The three control signals for the corresponding segment of the experiment.

As Fig. 2 shows, significant variations in, especially, u_c and u_h have been implemented. Consequently, the experiment should provide valuable data over a large range. To fit the uncertain parameters to the data, simulations are compared to measurements and a quadratic criterion defined by Eq. 6 is introduced.

$$V_f = \frac{1}{m} W^T \operatorname{diag}(X^T X) W \tag{6}$$

In Eq. 6, m is the number of observations, W is a weight vector that emphasizes the chosen variables, and X is a matrix with the simulation errors of interest as columns. The k-th row in X has the four columns $T_{\text{co},s}(k) - T_{\text{co}}(k)$, $T_{\text{hi},s}(k) - T_{\text{hi}}(k)$, $T_{\text{ho},s}(k) - T_{\text{ho}}(k)$ and $\dot{V}_{c,s}(k) - \dot{V}_c(k)$ and $1 \leq k \leq m$. Controlling T_{co} is the main objective in our application and motivates our choice of $W^T = \begin{pmatrix} \sqrt{10} & 1 & 1 & 1 \end{pmatrix}$.

In addition to the main experiment, experimental step changes in the control signals were performed to estimate delays. Based on visual inspection, delays from the control signals were estimated as 2, 2, and 6 seconds respectively. In other words, $u_c(t-2)$, $u_h(t-2)$ and $u_p(t-6)$ replaces u_c, u_h and u_p in the equations for the grey-box model. It can be noted that these estimates are somewhat arbitrary, e.g., is the delay from u_h to T_{co}, to T_{hi} or to T_{ho}? Maintaining the physical interpretation of the grey-box model, we choose delays to $\dot{V}_c$, $\dot{V}_h$ and P, respectively. Even so, variations between different step changes can be discerned. Moreover, for data-based models, other delays can be motivated.

Using constrained optimization, the estimates $\alpha = 0.30\,\text{kW}/(^\circ\text{Cm}^2)$, $k = 0.0033\,\text{kW}/^\circ\text{C}$, $C_c = 0.17\,\text{kJ}/^\circ\text{C}$, $C_h = 0.13\,\text{kJ}/^\circ\text{C}$ and $C_p = 5.9\,\text{kJ}/^\circ\text{C}$ are obtained. Simulations of T_{co} denoted $T_{\text{co},s}$ for the fitted model are illustrated along with measurements in Fig. 2.

3 Designing controllers

Traditionally, controller design is based on linear models of the dynamical systems. Rules-of-thumb approaches are based on simple models. With more detailed (linear) models, the design typically addresses one of, e.g., a desired stability margin, control performance as quantified by quadratic costs in control error and control signal activity, disturbance rejection, robustness by guaranteeing stable control under uncertainties, etc. Typically, these approaches require tools for differential and difference equations, linearizing equations, state-space descriptions and linear algebra, block diagrams, frequency analysis and Bode-diagrams, optimal control, etc. Accordingly, acquiring such skills forms a major focus of control engineering courses.

With increasingly accurate models, it appears that approaches based on linearized models do not take full advantage of available insights. Furthermore, maybe nonlinear approaches to control based on local linearity, such as gain scheduling, primarily are the result of adapting the design of controllers to traditional approaches?

As an alternative, mpc is not based on designing a static control law. Instead, control signal sequences are determined by optimizing simulations of a model to follow a desired trajectory over a predictive horizon. With new measurements, optimal control signal sequences are updated based on the available state. This approach has many attractive features, e.g., nonlinear models and constraints can easily be included. A disadvantage is that it may be difficult to determine the required computational complexity a priori, e.g., hard nonlinearities, bifurcations, etc., can render optimization unfeasible. Thus, simplifications (linearizing) can be needed to guarantee necessary computational efficiency.

In this paper, we instead combine the approach of designing a static control law with that of using simulations. With a static control law, the need for computational power in the real-time implementation is negligible and this is a key motivation behind our approach and a clear advantage compared to nominal mpc. The parameters in the controller are determined by optimization, i.e., similarly to mpc we formulate and minimize a criterion that quantifies differences in simulations of process values from desired setpoint trajecto-

ries taking control signal activity into consideration. The general criterion is similar to Eq. (6) and is given by

$$V_c = \frac{1}{m}(Q^T \operatorname{diag}(X_c^T X_c)Q + R^T \operatorname{diag}(\Delta u^T \Delta u)R \quad (7)$$

where X_c is given by differences between setpoint (r) and simulated process value (y) and its k-th row is given by

$$X_c(k) = \begin{pmatrix} r_1(k) - y_1(k) & r_2(k) - y_2(k) & \cdots \end{pmatrix} \quad (8)$$

In our quadratic criterion, control signal activity is quantified by the change in control signal, $\Delta u = u(k+1) - u(k)$, but u can naturally be used directly if more suitable for the application at hand.

The idea of simulation-based controller design is not new, e.g., neural network controllers can often be seen as a version of this approach. But the idea and criterion has to the best of our knowledge not been as explicitly discussed as we try in this paper. Furthermore, we also discuss how to incorporate feedback systems analysis within our framework.

For our specific case study, the focus of the present study is to explore possibilities for controlling T_{co} using u_h. In Eq. (7), $X_c = T_{\text{co,sp}} - T_{\text{co,s}}$ and $\Delta u = \Delta u_h$ accordingly, and with the heuristic choices of $Q = 20\sqrt{2}$ and $R = 1$ based on evaluations of simulations. For the other control variables, the flow on the cold side, $\dot{V}_c$, with the pump on the cold side, u_c, and the temperature entering the heat exchanger on the hot side, T_{hi}, with u_p, we use fixed PI-controllers.

For the PI-controllers, we use the velocity equation

$$u_i(k) = u_i(k-1) + K_i \left(\left(1 + \frac{T_s}{T_{I,i}} \right) e_i(k) - e_i(k-1) \right) \quad (9)$$

with control signals constrained between 0 and 100%, $u_i(k) = \min(\max(u_i(k), 0), 100)$. In Eq. (9), K is the (proportional) gain, T_I the integration time, T_s the sampling period ($T_s = 1$ s in all simulations and experiments) and $e_i(k) = r_i(k) - y_i(k)$ is the control error, i.e., the difference between setpoint and process value. The subscript i is either c for the pump on the cold side, h for the pump on the hot side or p for the command to the power in the heater. The PI-controller for u_c has $K_c = 0.01$ %/(ml/min) and $T_{I,c} = 0.5$ s and u_p has $K_p = 20$ %/°C and $T_{I,p} = 100$ s[1].

To explore realistic challenges, the setpoint for $\dot{V}_c$ changes from 150 ml/min to 220 ml/min at $t = 1000$ s as a ramp stretching over 5 seconds and then to 100 ml/min at $t = 2000$ s as a ramp stretching over 10 seconds. These setpoint changes can be seen in Fig. 3. The setpoint for T_{co} changes from 35 °C to 40 °C at $t = 500$ s, back to 35 °C at $t = 1500$ s and to 40 °C at $t = 2500$ s. All changes are ramp-shaped and stretch over 10 seconds. These setpoint changes can be seen in Figs. 4 and 6. The setpoint for T_{hi} is kept constant at 55 °C.

[1] These choices are at least not intentionally biased for our study as they are arbitrarily chosen from submitted student assignments.

3.1 PI control

Minimizing Eq. (7) for the simulated response to the setpoint changes regarding K_h and $T_{I,h}$ gives the results $K_h = 4.7$ %/°C and $T_{I,h} = 23$ s. Corresponding simulations are illustrated in the left panels of Figs. 3–5. For reference and comparison, the value for our criterion of Eq. (7) is $V_{c,e} = 510$. In Figs. 3–5, experimental results using the same setpoint changes are illustrated in the right panels. In addition to the presence of measurement noise, some differences can be noted. In the simulations, T_{co} does not reach the setpoint after the change in $\dot{V}_c$ at $t = 1500$ and u_h saturates at 100%. In the experiment, this disturbance in T_{co} is significantly smaller and compensated for by the controller with u_h just under 70%. This discrepancy between model and experiment could be explained by a higher efficiency of the heat exchanger at higher flows, i.e., α could better be described as a function of flows and, possibly, temperatures. For the control, however, it appears as if the well-known strength of integral action in the controller is robust against such low-frequency modelling errors. Perhaps more interesting, the disturbance in T_{co} at $t = 2000$ due to the change in $\dot{V}_c$ causes slowly converging oscillations, suggesting a bifurcation in the simulated closed-loop system that could warrant further study. As an advantage in practice, we note that control of T_{co} is similar but better in experiments than in simulations.

For the control of T_{hi} much larger high-frequency variation in both T_{hi} and u_p can be noted in the experiment compared to simulations. Although the control on a general level is similar in simulation and experiment and works well since high-frequency variations in u_p are not a matter of concern, this discrepancy will be briefly discussed in Section 4.

For the experiment, the value $V_{c,e} = 140$ is obtained. This significantly lower value despite measurement noise compared to the simulation can partly be explained by the simulated T_{co} not reaching the setpoint for $1500 \leq t \leq 2000$ as noted before.

3.2 Nonlinear PI control

In addition to enabling the use of detailed and, possibly, opaque models, one of the points behind using simulations for controller design is the possibility for exploring alternative, nonlinear controllers. As the left panels in Figs. 4–5 clearly reveal, the model exhibits challenging nonlinear characteristics. An intuitive solution could therefore be to consider a nonlinear PI-controller. Instead of gain-scheduling with

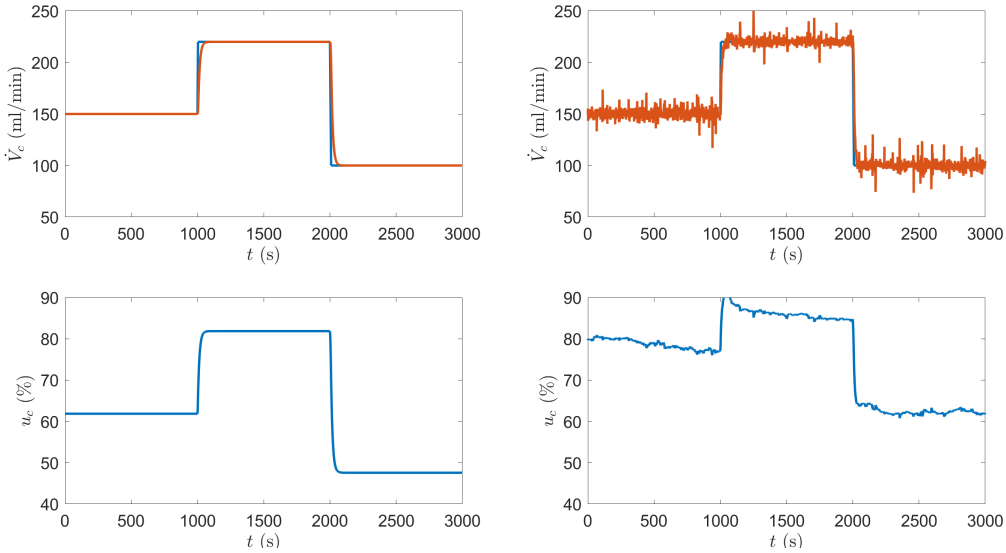


Figure 3. Upper panels: Setpoint changes in of $\dot{V}_c$ (blue) and simulated control (red) in the left panel, and the corresponding experiment to the right. Lower panels: Simulated control signal u_c (left)

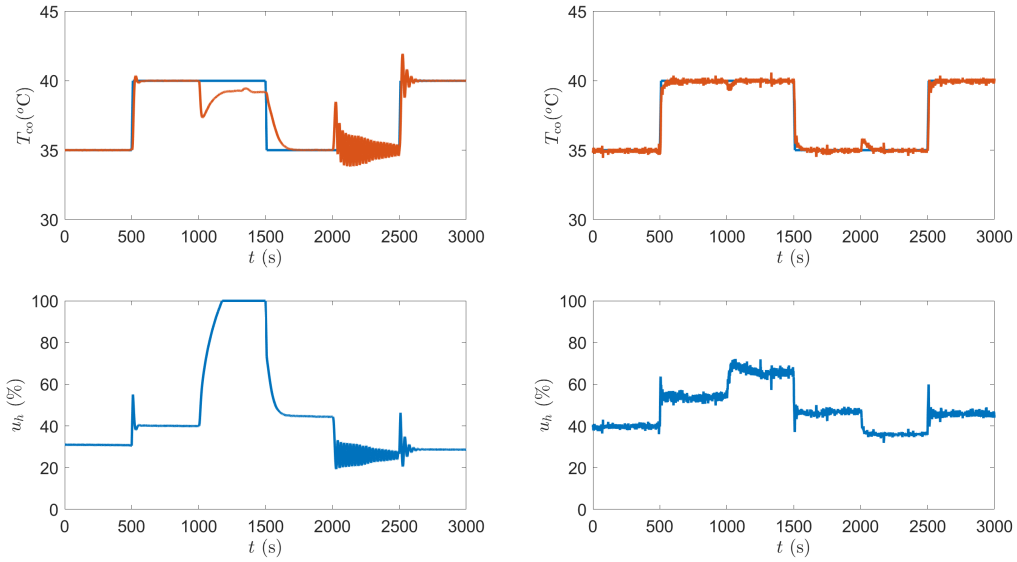


Figure 4. Upper panels: Setpoint changes in of T_{co} (blue) and simulated control (red) in the left panel, and the corresponding experiment to the right. Lower panels: Simulated control signal u_h (left) and corresponding experiment (right).

different linear PI-controllers depending on a parameter (typically u, y or r), we explore linear dependencies in the gain and integration time. Specifically, in place of K_h and, $T_{I,h}$ we use $K_h + au_h(k-1)$ and $T_{I,h} + bu_h(k-1)$.

Minimizing Eq. (7) for the simulated response to the setpoint changes regarding K_h, a, $T_{I,h}$ and b gives the results $K_h = 0.1$ %/°C, $a = 46$ 1/°C, $T_{I,h} = 0.18$ s and $b = -0.46$ s/%. Compared to the linear PI-controller of Eq. (9) this corresponds to a range of values for K_h between 3.5 %/°C and 18 %/°C and for $T_{I,h}$ between 0.14 s and 38 s. Simulations yield the value $V_{c,e} = 240$, i.e., a significant improvement over the linear case.

The experiment with the nonlinear PI-controller yields the value $V_{c,e} = 52$. Compared to the experiment with the linear controller, the performance is numerically clearly better. Visual inspection of Figs. 6–7 reveal that setpoint tracking is faster and disturbance rejection better than in the linear case, at the cost of higher activity in the control signal u_h. In summary, the results are very encouraging and a significant improvement in control quality can be noted.

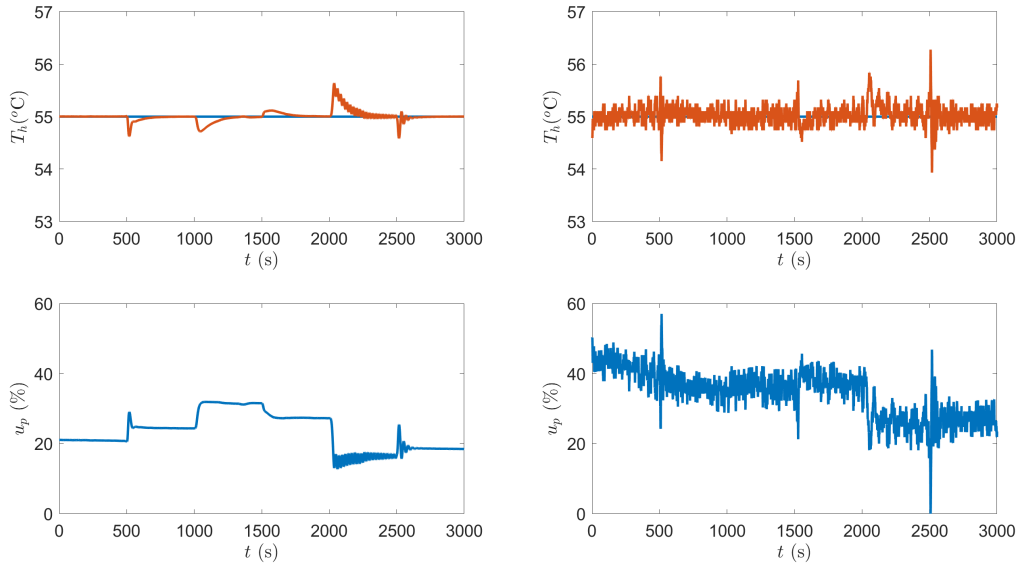


Figure 5. Upper panels: Setpoint for T_{hi} (blue) and simulated control (red) in the left panel, and the corresponding experiment to the right. Lower panels: Simulated control signal u_p (left) and corresponding experiment (right).

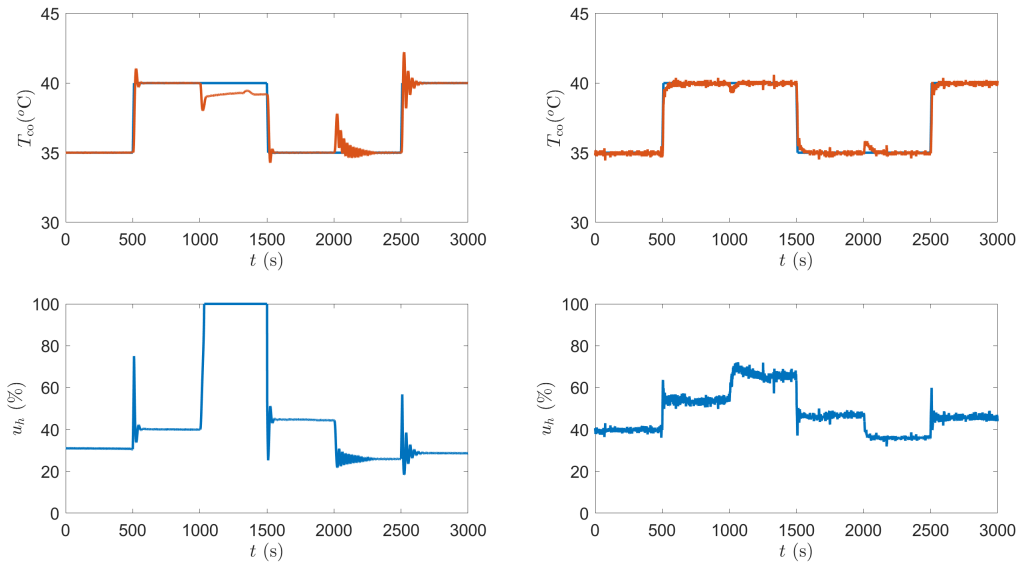


Figure 6. Same as in Fig. 4 but for the nonlinear PI-controller.

4 Control system analysis

It is commonly known that control design methods "focus on one or two aspects of the [control] problem, and the control-system designer then has to check that the other requirements are also satisfied" (Åström & Wittenmark, 1997). Using a block diagram and transfer function notation with s as the Laplace variable, a feedback system following the structure we use is depicted in Fig. 8.

Correspondingly,

$$Y(s) = \frac{G_r G_p}{1+G_r G_p} R(s) + \frac{G_p}{1+G_r G_p} W_1(s) + \frac{1}{1+G_r G_p} W_2(s) \quad (10)$$

and

$$U(s) = \frac{G_r}{1+G_r G_p} R(s) - \frac{G_r G_p}{1+G_r G_p} W_1(s) - \frac{G_r}{1+G_r G_p} W_2(s) \quad (11)$$

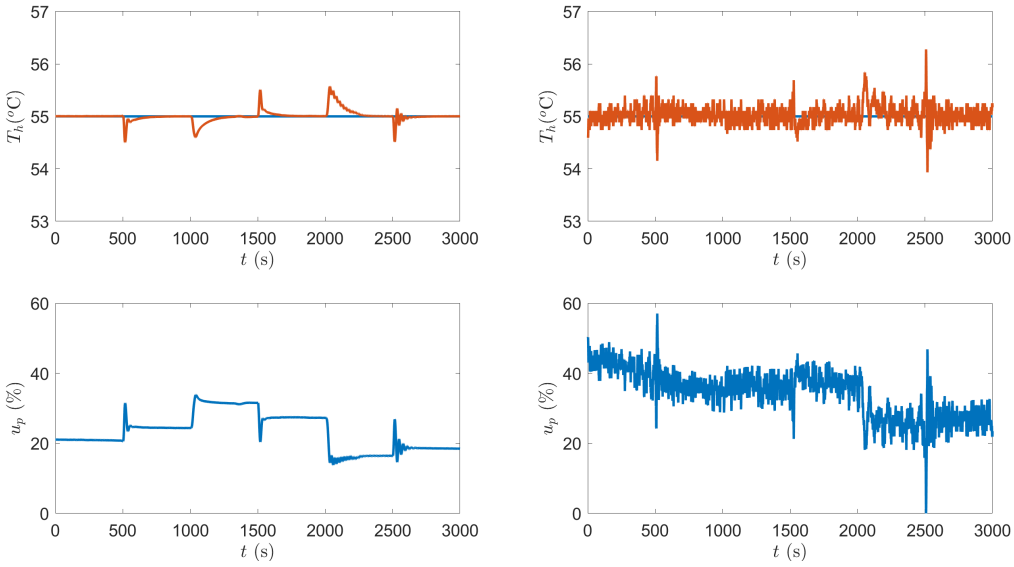


Figure 7. Same as in Fig. 5 but for the nonlinear PI-controller.

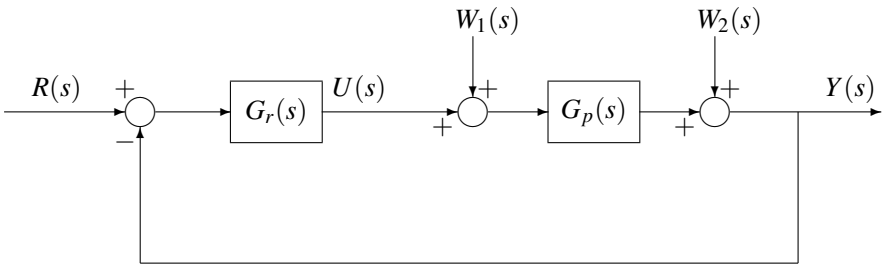


Figure 8. Block diagram for a process $G_p(s)$ controlled by (negative) feedback. The diagram includes setpoint $R(s)$, control signal $U(s)$, process disturbance $W_1(s)$, measurement noise $W_2(s)$, process value $Y(s)$ and controller $G_r(s)$.

As these equations indicate, the control problem can, in turn, often be analyzed (in the frequency domain) by addressing the "Gang Of Four" (Åström & Murray, 2021), i.e, the four transfer functions

$$\frac{G_r G_p}{1+G_r G_p} \tag{12}$$

known as the complementary sensitivity function,

$$\frac{G_p}{1+G_r G_p} \tag{13}$$

the load sensitivity function,

$$\frac{1}{1+G_r G_p} \tag{14}$$

the sensitivity function and

$$\frac{G_r}{1+G_r G_p} \tag{15}$$

the noise sensitivity function. For linear systems, it is illustrative to analyze the closed-loop system with plots of (the gains of) these transfer functions as a function of frequency. In principle, simulations using sinusoidal functions of different frequencies for, e.g., $w_2(t)$ and recording the corresponding amplitudes of $y(t)$ and $u(t)$ could provide numerical estimates of these gains. However, nonlinear systems can exhibit, e.g., frequency spreading, i.e., a single frequency in $w_2(t)$ can result in several frequencies in $y(t)$ and $u(t)$. This phenomenon can further be amplitude-dependent and, for multivariable systems, the principle of superposition is not necessarily applicable. In addition, the behavior can depend on the region of operation.

In summary, these characteristics render a full exploration of our case study cumbersome at the very least. We are still grasping at how the abundance of combinations of different variables could be illustrated. As an initial exploration, and motivated by the challenges the experiments indicate, we separately explore how sinusoidal measurement noise in T_{co} and T_{hi} for different frequencies affect T_{co}, u_h, T_{hi} and u_p, respectively. Constant setpoints are used, 37.5°C for T_{co} and

160 ml/min for $\dot{V}_c$. The amplitude of the measurement noise was 1°C and possible frequency spreading was neglected, i.e., only amplitudes of corresponding oscillations were recorded. The corresponding results are illustrated in Figs. 9–10.

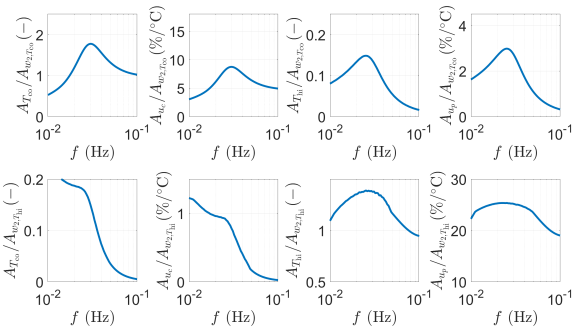


Figure 9. Amplitude of variations as a function of frequency resulting from a sinusoidal measurement disturbance on T_{co} (upper panels) with amplitude 1°. Left to right, amplitudes of T_{co}, u_c, T_{hi} and u_h. Lower panels: Same as upper panels but for T_{hi} with amplitude 1°.

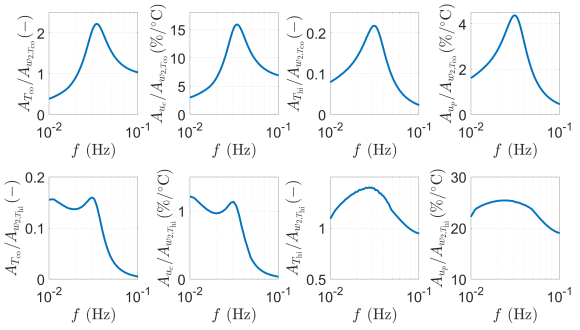


Figure 10. Same as in Fig. 9 but for the nonlinear PI-controller.

As the figures reveal, similar "sensitivity functions" are obtained for the linear and nonlinear PI-controllers, indicating a (limited) insensitivity to process variations. Based on this limited analysis, the main source of possible concern is in the "noise sensitivity functions": variations in T_{hi} can be amplified to the control signal u_p by a factor of 25 and to u_h by a factor of 3 using the linear PI-controller and a factor of 4 using the nonlinear PI. Variations in T_{co}, in turn, can be amplified to the control signal u_h by a factor of 9 (linear PI) and by 16 (nonlinear PI).

This analysis can explain the control signal activity observed in u_h (lower right panels of both Fig. 5 and Fig. 7) as well as the difference between control signal activity observed in u_p (lower right panels of Fig. 4 and Fig. 6). Most likely, the high gain of 25 is due to the heuristically designed PI-controller for controlling T_{hi} using u_p and it can be noted that the gain $K_p = 20$ %/°C was used. Still, a comparative study of different approaches for controller design, possibly including simulation-based loop-shaping, could help to shed light on possible weaknesses in the simulation-based approach. Moreover, an interesting possibility could be to include, e.g., the maximum gains of noise sensitivity functions in the criterion minimized for designing controllers, Eq. (7).

5 Conclusions and future work

In this paper, we explored simulation-based controller design. Compared to approaches based on linearized models, this enabled us to explore and tune alternative nonlinear controllers. Moreover, the approach does not rely on models that are translucent and easily linearized. For our challenging case study on a multivariable, nonlinear heating system, the simulations were promising and experimental results above expectations. Further experiments, study, analysis, and comparisons to other promising frameworks such as reinforcement learning or mpc will be pursued in order to explore the general applicability of the presented approach.

References

Åström, K. J., & Murray, R. M. (2021). *Feedback systems: An introduction for scientists and engineers* (2nd ed.). Princeton University Press.

Åström, K. J., & Wittenmark, B. (1997). *Computer controlled systems: Theory and design* (3rd ed.). New Jersey: Prentice-Hall.

Dettù, F., Formentin, S., & Savaresi, S. M. (2023). The twin-in-the-loop approach for vehicle dynamics control. *arXiv preprint arXiv:2209.02263*.

Espinosa-Leal, L., Chapman, A., & Westerlund, M. (2020). Autonomous industrial management via reinforcement learning. *Journal of intelligent & Fuzzy systems*, *39*(6), 8427–8439.

Hewing, L., Wabersich, K. P., Menner, M., & Zeilinger, M. N. (2020). Learning-based model predictive control: Toward safe learning in control. *Annual Review of Control, Robotics, and Autonomous Systems*, *3*, 269–296.

IPCC. (2023). Summary for policymakers [Book Section]. In H. Lee & J. Romero (Eds.), *Climate change 2023: Mitigation of climate change. contribution of working groups i, ii, iii to the sixth assessment report of the intergovernmental panel on climate change.* Geneva Switzerland: IPCC, in press.

Pintelon, R., & Schoukens, J. (2012). *System identification—a frequency domain approach* (2nd ed.). New Jersey: John Wiley & Sons.

Multimodal sensor suite for identification of flow regimes and estimation of phase fractions and velocities – Machine learning algorithms in multiphase flow metering and control

Noorain Syed Kazmi [a,*], Ru Yan [a], Saba Mylvaganam [a], Håkon Viumdal [a]

[a] *University of South-Eastern Norway, Faculty of Technology, Natural Sciences and Maritime Sciences, Dept. EE, IT and Cybernetics, Kjølnes Ring 56, 3918 Porsgrunn, Norway*

* 238758@usn.no

Abstract

Multiphase flow metering is a challenging task because of the complexity of multiphase flow. In this paper, non-intrusive multiphase flow metering techniques, including machine learning (ML) / artificial intelligence models for the identification of flow regimes and estimation of flow parameters of a two-phase flow in a horizontal pipe are proposed that use data from Electrical Capacitance Tomography (ECT) and conventional measurements such as differential pressure in the pipe. The flow regimes are classified into five types, namely plug, slug, annular, wavy and stratified. Two-phase air/water flow experimental data from ECT are collected by running extensive experiments using the horizontal section of the multiphase flow rig at the University of South-Eastern Norway (USN). Exploratory data analysis (EDA) is performed on these data to extract features for use in classification and regression algorithms. Time series of normalized capacitance data from ECT sensors are used to classify flow regimes and identify flow parameters. ML techniques of Artificial Neural Network, Support Vector Machine (SVM), K-Nearest Neighbors (KNN) and Decision Tree (DT) are used to classify flow regimes by using features extracted from ECT data. The cross-correlation technique is used to estimate flow velocity using data from a twin-plane ECT module. ML regression techniques are used to estimate phase fractions. Fusing data from differential pressure sensors enhances the flow regime classification. An overall system performance is given with suggestions for designing dedicated control algorithms for actuators used in multiphase flow control.

1. Introduction

In fluid mechanics, multiphase flow is the flow of two or more phases of matter in a pipe. Multiphase flow is a complex phenomenon. Two-phase flow is a flow where two phases out of solid, liquid and gas phases are observed simultaneously in a pipeline. Gas/Solid is prevalent in pneumatic conveyors, dust collectors, fluidized beds, heterogeneous reactors and metallized propellant rockets. Gas/liquid flow can be seen in atomizers, scrubbers, dryers and combustors. Liquid/liquid droplet flow is observed in extraction, homogenizing and emulsifying. Liquid/solid is present in flotation and sedimentation (Soo, 1990).

The geometric distribution of constituent phases in a multiphase flow is known as flow regime or pattern (Tan and Dong, 2023). There are various types of flow regimes. Slug, plug, stratified, annular, wavy, bubble, etc., are common and well-known flow regimes observed in multiphase flow (Vohr, 1960). Flow regimes depend on the orientation of pipe and direction of flow. The density of phases, viscosity of phases and mass flow rates of phases also greatly affect the creation of flow regimes (Alssayh et al., 2013). Operating pressure, temperature, valves and bends have a direct effect on the flow regimes (Hansen et al., 2019). Classification of flow regimes in a two-phase flow pipeline is a major challenge in the field of flow analysis (Pereyra et al., 2012). Flow regimes can be classified subjectively through graphics or by employing the probability density function of pressure or void fractions signals from sensors (Almalki and Ahmed, 2020; Godfrey Nnabuife et al., 2021). Flow regimes has direct effect on the measurement of flow velocities, phase fractions and other parameters (Godfrey Nnabuife et al., 2021). Some of the flow regimes observed in horizontal gas/liquid multiphase flows are described below:

- Stratified

When gas and liquid flow rates are low, stratified flow is observed. It is applicable in horizontal flow direction. There is no mixing of the two phases and the liquid phase remains as a film at the lower portion of the pipe (Liné and Fabre, 2011).

- Wavy

At higher gas flow rates, the stratified flow converts to wavy flow in which ripples or waves are observed on the top of the liquid layer. It appears like waves in a sea. (Jayanti, 2011)

- Annular

At higher gas velocity, wavy flow converts to annular flow in which liquid flows at the periphery of the pipe while gas flows at the center of the pipe. (Zeigarnik, 2011)

- Plug / Elongated Bubble

Plug is a kind of flow pattern in which large bubbles of gas float on the top surface of the liquid phase spanning a large area in the pipe. The gas phase is dispersed in the liquid phase. (Vohr, 1960)

- Slug

Slug flow is intermittent flow in which slugs of liquid with dispersed bubbles flow along with large gas pockets. The flow is always unsteady. The bubble is in the shape of a bullet. This flow alternates between high liquid fraction and high gas fraction. (Vohr, 1960)

(Wang and Zhang, 2009) use an ECT for identifying flow regimes by applying Support Vector Machine (SVM) to classify flow regimes and Principal Component Analysis (PCA) to optimize the inputs to the SVM model. (Ameran *et al*., 2015) discuss velocity measurement of two-phase flow through ECT using cross-correlation techniques. A study of flow velocity and phase concentrations of horizontal two-phase flow is presented by (Stavland *et al.*, 2021), employing a dual-plane ECT with gamma-ray tomography to measure volumetric flow rates of the phases, achieving an accuracy of ±10%.

The results presented in this paper are developed during the master thesis's work of (Noorain Syed Kazmi, 2023).

2. System Description

A multiphase flow rig is at the University of South-Eastern Norway (USN), Campus Porsgrunn. This rig is equipped with facilities for multiphase flow studies–using water, air, and mineral oil through a horizontal pipe. The pipe can be tilted by ±10º with respect to the horizontal surface. The operational limit of mass flow rate for air is 5 kg/min, whereas for liquid is 150 kg/min. By injecting various combinations of air, water and oil mass flow rates, different flow regimes can be generated and visually inspected through the Plexiglass transparent section, as shown in Fig. 1. A simplified piping and instrumentation diagram (P&ID) of the flow rig is shown in Fig. 2. Some important parameters of the rig are given in Tab. 1.

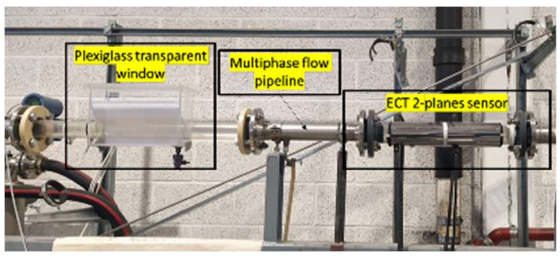


Figure 1: The rig setup (partly) at USN, Porsgrunn.

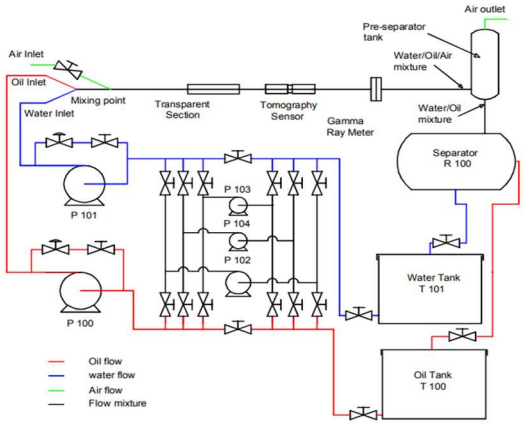


Figure 2: P&ID of the rig at USN, Porsgrunn

Table 1: Some parameters of the rig at USN, Porsgrunn

USN Flow Rig	
Inside pipe diameter	56 mm
Outside pipe diameter	60 mm
Length of pipe	15 m
Water density	996 kg/m3 at 27ºC and 1bar
Air density	1 kg/ m3
Oil density	790 kg/m3
Water viscosity	0.00102 Pa-s (20ºC)
Oil viscosity	0.00164 Pa-s (25ºC)

As depicted in Fig. 1, a TOMOFLOW TFLR5000 dual-plane ECT system from Process Tomography Limited is equipped on the rig. The ECT system can measure the flow parameters of an uneven two-phase flow when the constituents have dielectric properties (Process Tomography Limited, 2011). In this rig, air, oil and water permittivities are 1, 2.7 and 80, respectively (Dupré et al., 2017).

PDT120 and PDT121 are the differential pressure meters mounted on the rig, as shown in Fig. 3. PDT120 measures the differential pressure across a span of 10.22m in the pipe, and PDT121 captures the differential pressure across a shorter distance of 5.38m within the same pipe. In addition, the inlet air flow rate and air pressure are measured separately by a flow transmitter FT131 and a pressure transmitter PT131.

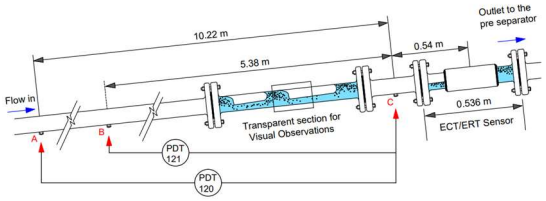


Figure 3: Measurement areas of differential pressure meters on the flow rig at USN, Porsgrunn (Dupré et al., 2017)

3. Electrical Capacitance Tomography

ECT is a non-invasive, non-radioactive flow sensing method that measures the spatial distribution of dielectric materials within a pipe using capacitance readings from peripheral electrodes (Process Tomography Limited, 2011; Saied and Meribout,

2016). The capacitances can also be converted into images for visual depiction of the flow.
Generally used in two-phase liquid/gas flow, ECT measures permittivity distribution inside a pipe at two cross-sections of a pipe that can give the velocity profile of flow while also providing volume ratio or phase fraction data. ECT is most effective when working with fluids that have low electric conductivity and variable permittivity (Process Tomography Limited, 2011). The working principle of an ECT (one plane) is shown in Fig. 4.

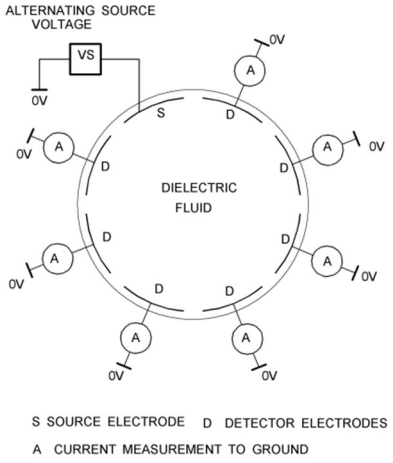


Figure 4: Principal of ECT with 8-electrodes (Process Tomography Limited, 2011)

An alternating voltage is applied between the source and the ground. The source is charged to one electrode. Currents, in direct proportion to the capacitance, are subsequently measured at all remaining electrodes. Within a single measurement frame, the currents/ capacitances between each pair of electrodes are measured. For an *N*-electrodes ECT plane, there are in total $M = N(N-1)/2$ unique capacitance values per measurement frame (Process Tomography Limited, 2011). The measured capacitances can be normalized by using Eq. 1.

$$C_n = \frac{C_m - C_L}{C_h - C_L} \qquad (1)$$

In Eq. 1, C_m is the inter-electrode raw capacitance. C_L is the capacitance when the pipe is full with lower permittivity material such as air. C_h is the capacitance when the pipe is full with higher permittivity material such as water. C_n is the normalized capacitance. C_n is dimensionless and normalized, making it suitable as input for mathematical operations and algorithms.

4. Experiments

Based on the flow conditions outlined in Fig. 5, 45 of two-phase air and water experiments are carried out on the flow rig using the ECT system. Conventional measurements such as differential pressure, temperature and mass flow rate were also recorded during each experiment.
The flow regimes indicated in Fig. 5 are validated via visual inspection throughout the experiments. Fig. 6 displays the active experimental area on the flow regime map. The lowest flow rates for air and water are 0.07 kg/min and 2 kg/min, respectively, while the highest flow rates for both mediums reach 5 kg/min and 77 kg/min. Fig. 7 illustrates the setup of the sensor array, comprising of 8 electrodes, around the pipe. Tab. 2 provides the parameters setup in the ECT system during the experiments.

Test Matrix		Water (kg/min)					
		2	3	4	75	76	77
Air (kg/min)	0.07				P	P	P
	0.09				P	P	P
	0.1	ST	ST	ST			
	0.11				P	P	P
	0.13				P	P	P
	0.15	ST	ST	ST	P	P	P
	0.3				S	S	S
	0.4				S	S	S
	0.5	ST	ST	ST	S	S	S
	1	W	W	W			
	2.5	W	W	W			
	4	A	A	A			
	5	A	A	A			
ST: Stratified; W: Wavy; A: Annular; P: Plug; S: Slug							

Figure 5: Test matrix for two-phase flow with varying velocities of water and air (in kg/min) generating different flow regimes

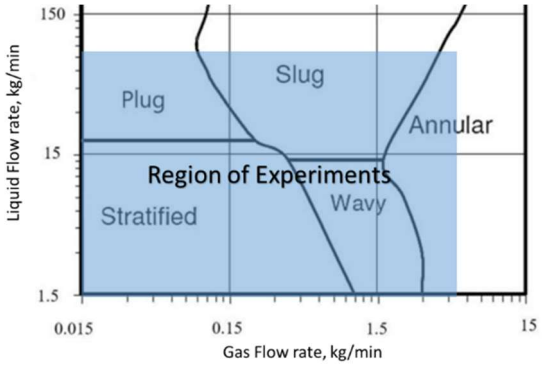


Figure 6: Active region (blue area) of experiments on a flow regime map, based on Mandani et al..

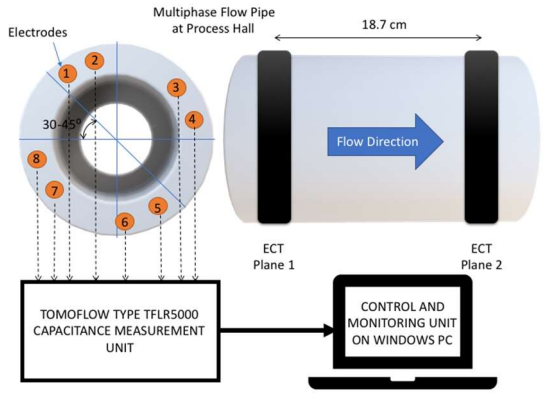


Figure 7: Setup of ECT electrodes and planes on the pipe and their connection to TOMOFLOW TFLR5000 system

Table 2: HW/SW related parameters used in the ECT module for the experiments

Parameter	Value
Frames per second per plane	100
Sampling interval per plane (ms)	10
Number of planes	2
Number of electrodes per plane	8
Number of capacitances per measured frame	28
Logging duration per experiment (s)	30

For each experiment, 3000 frames of ECT data were collected. For example, one of the experiments for Annular flow regime was conducted by simulating an Annular flow in the USN flow pipeline by using the matrix of Fig. 5. This experiment generated 3000 frames of normalized capacitances from the eight electrodes. This batch of frames was labeled as Annular to be used for supervised machine learning algorithms. Each frame consists of 28 capacitance values. For 45 experiments, a total of 135000 frames of capacitance data are collected. These capacitance data were normalized before using them in classification and regression algorithms. Each frame was flattened to 28 columns of normalized capacitances with the observed flow regime in the 29th column of the flattened file enabling 135000 rows as inputs to machine learning models.

5. Methods

This paper defines CXY as the normalized capacitance between electrodes X and Y, with C12, for instance, denoting the normalized capacitance between electrodes 1 and 2. The electrode counts are given in Fig.7. The classification and regression models are developed in MATLAB R2020b for this paper.

5.1 Flow Regime Identification

Flow regime identification utilizes 28 normalized capacitances from one ECT data frame as features/inputs. The associated flow regime types act as labels/outputs in machine learning (ML) classification algorithms, as illustrated in Fig. 8.

ML algorithms of Decision tree (DT), K-Nearest Neighbors (KNN), SVM and Feedforward Neural Networks (FNN) are used as flow regime classification algorithms. Classification Learner App in MATLAB is used to develop the flow regime classification algorithms of DT, KNN and SVM. The Neural Network Pattern Recognition App in MATLAB is used to develop the flow regime classification FNN algorithm.

In pursuit of enhancing model performance, another model incorporating both ECT data and differential pressure data from sensors PDT120, PDT121, and PT131 is also developed by implementing a sensor fusion method, as illustrated in Fig. 9

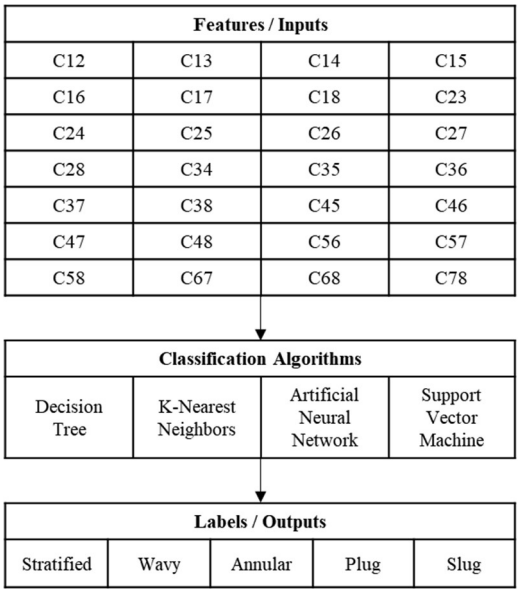


Figure 8: Model for flow regime classification using normalized capacitances as features in ML algorithms

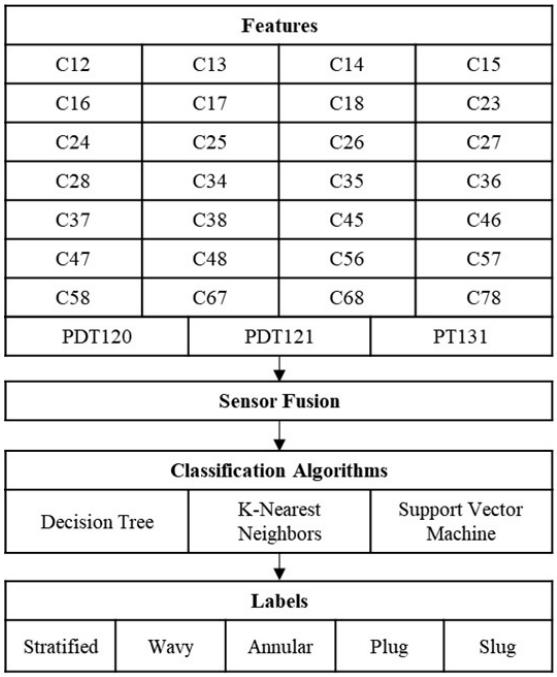


Figure 9: Model for flow regime classification using sensor fusion concept

5.2 Flow Velocity Estimation

To estimate the flow velocity, a cross-correlation analysis is performed using frames from both planes in a dual plane of ECT. The normalized capacitances at these two planes are cross-correlated to find the peak correlation lag between them. For instance, the series of C12 at plane 1 is correlated with the series of C12 at plane 2 in the ECT sensor. The peak corresponds to the degree of similarity between the two capacitances. This model specifically considers Annular, Plug, and Slug flow regimes due to their dynamic flow characteristics. The distance between the two planes mounted on the rig is 0.187 m. The flow velocity is calculated by dividing this distance by the lag time, where each lag time is 10ms.

5.3 Volume Ratio Estimation

The volume ratio for each phase is estimated by considering two differential pressure data from PDT120 and PDT121, and the inlet air flow rate data from FT131 as inputs to the regression model. The inverse volume ratio data obtained from ECT experiments is considered the model training targets, as explained in Fig. 10.

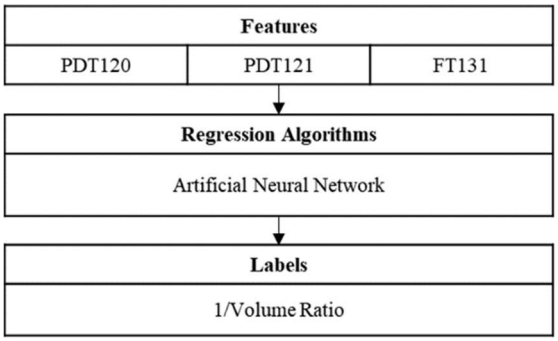


Figure 10: Model for volume ratio estimation with two pressures and flow measurement as inputs.

FNN is used as a volume ratio estimation algorithm. The trained FNN model features a single hidden layer containing 10 neurons.

6. Results

6.1 Flow Regime Identification

Using the model illustrated in Fig. 8, the flow regime classification neural network examines a total of 135000 samples with an evenly distributed array of flow regimes. For training and testing of the FNN model, these samples are divided into training, validation, and testing datasets in a 70:15:15 ratio. The hidden layer in the FNN employs a *tansig* activation function. The performance of the neural network, as seen in Fig. 11, indicates an overall accuracy of 96.5%.

KNN, SVM and DT algorithms are utilized for training flow regime classification models. Half of the data is reserved for validation purposes. Tab. 3 presents the overall validation accuracy achieved by these algorithms in classifying flow regimes.

KNN gives the highest accuracy, while SVM has the lowest accuracy among the three. The confusion matrix of KNN is shown in Fig. 12. This model will perform well when the flow is in the region of the training data as per Fig. 5. The data from transition zones of the flow regime matrix was not used to train this model.

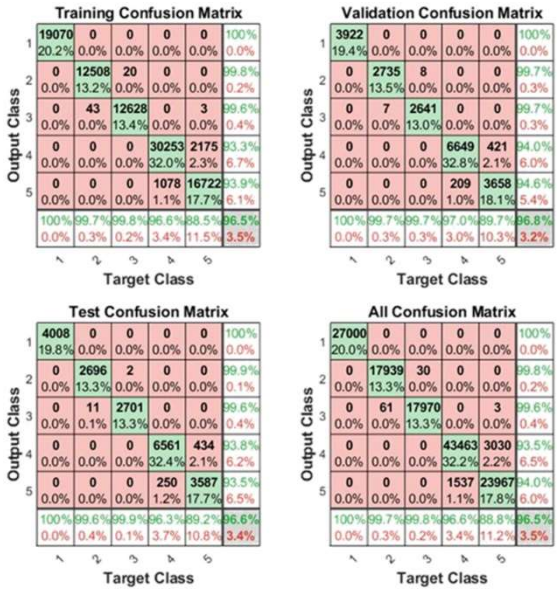


Figure 11: Performance of the FNN model for flow regime classification with 28 normalized capacitances as inputs and flow regime (1-Stratified, 2-Wavy, 3-Annular, 4-Plug, 5-Slug) as output.

Table 3: Comparison of accuracy from various flow regime identification models using 28 normalized capacitances as inputs and flow regime as output

ML algorithms	Overall accuracy (%)
KNN (Fine)	98.7
DT (Fine)	96.6
SVM (Linear)	94.7

Figure 12: Confusion Matrix of KNN (Validation) for flow regime identification with 28 normalized capacitances as inputs and flow regime as output. The corresponding sample amounts are represented by percentages and detailed in the parentheses below.

In the sensor fusion-based model, illustrated in Fig. 9, pressure and differential pressure meter signals merge with normalized capacitances to serve as features/inputs. Training of the models continues to use KNN, SVM, and DT algorithms. However, due to differing data sampling frequencies (PDT and PT sampled at 20 Hz, while ECT at 100 frames per second per plane), the total sample count is reduced for synchronization. The overall validation accuracy from these algorithms to classify flow regimes is given in Tab. 4.

Table 4: Comparison of accuracies achieved with various algorithms for flow regime identification using 28 normalized capacitances and the three pressure signals as inputs and flow regime as output

ML algorithms	Overall accuracy (%)
KNN (Fine)	98.6
DT (Fine)	98.6
SVM (Linear)	99

SVM gives the highest accuracy. The confusion matrix of SVM is shown in Fig. 13.

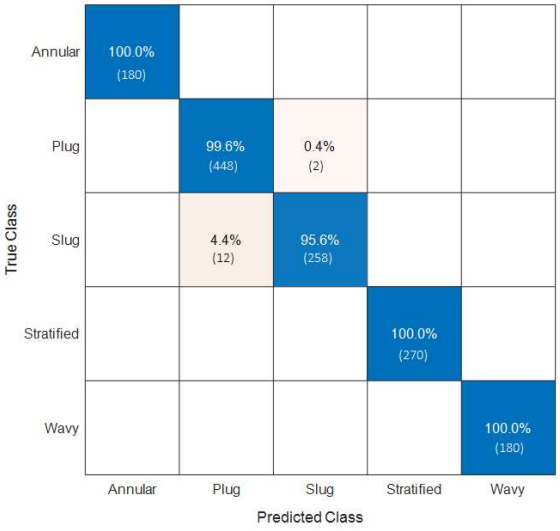


Figure 13: Confusion Matrix of SVM (Validation) for flow regime identification with 28 normalized capacitances and 3 pressure signals as inputs and flow regime as output. The corresponding sample amounts are represented by percentages and detailed in the parentheses below.

6.2. Flow Velocity Estimation

The results of estimating flow velocity are presented in the subsequent subsections, featuring two cases for each flow regime: annular, plug and slug.

6.2.1. Annular

- Case 1: Water - 2 kg/min, Air – 4 kg/min

Higher lags are disregarded as they likely arise from random fluctuations. Thus, from Tab. 5, there are 10 of found lags that are disregarded as seen marked by the orange-colored rectangles. Therefore, the domain cross-correlation lags for the remaining capacitances are at around -13, ignoring the “0” lags.

8 lags are corresponding to approximately 120 to 130ms time. Therefore, the flow velocity is estimated to be 0.187/0.13 = 1.43 m/s for this experiment.

Table 5: Cross-correlation for Water - 2 kg/min, Air – 4 kg/min (*Disregarded lags marked by the orange-colored rectangles*)

Normalized Capacitance	Peak Correlation Lag	Normalized Capacitance	Peak Correlation Lag
C12	0	C35	0
C13	0	C36	-12
C14	0	C37	-1664
C15	0	C38	-2992
C16	-13	C45	0
C17	-13	C46	-13
C18	-13	C47	-2992
C23	-175	C48	-2992
C24	0	C56	-35
C25	0	C57	-2787
C26	-12	C58	2997
C27	-2993	C67	-13
C28	-2963	C68	-13
C34	0	C78	0

- Case 2: Water - 2 kg/min, Air – 5 kg/min

Higher lags are disregarded as they likely arise from random fluctuations. Thus, from Tab. 6, there are 9 of found lags that are disregarded. Therefore, the domain cross-correlation lags for the remaining capacitances are at around -10, ignoring the "0,-1" lags. 10 lags are corresponding to approximately 60 to 100ms time. Therefore, the flow velocity is estimated to be 0.187/0.09 = 2.07 m/s for this experiment.

Table 6: Cross-correlation for Water - 2 kg/min, Air – 5 kg/min

Normalized Capacitance	Peak Correlation Lag	Normalized Capacitance	Peak Correlation Lag
C12	0	C35	0
C13	0	C36	-9
C14	0	C37	-2994
C15	-6	C38	-2997
C16	-10	C45	-1
C17	-10	C46	2997
C18	-10	C47	-2997
C23	-2997	C48	-2996
C24	0	C56	-9
C25	0	C57	-10
C26	2860	C58	-9
C27	-2994	C67	-10
C28	2997	C68	-10
C12	0	C35	0

6.2.2. Plug

- Case 1: Water - 76 kg/min, Air – 0.11 kg/min

Tab. 7 suggests the domain cross-correlation lags for the capacitances are at around -12. 28 lags are corresponding to approximately 100 to 130ms time. Hence, this experiment estimates a flow velocity of 0.187/0.12 = 1.55 m/s for this experiment.

Table 7: Cross-correlation for Water - 76 kg/min, Air – 0.11 kg/min

Normalized Capacitance	Peak Correlation Lag	Normalized Capacitance	Peak Correlation Lag
C12	-13	C35	-12
C13	-12	C36	-12
C14	-10	C37	-12
C15	-12	C38	-12
C16	-11	C45	-10
C17	-11	C46	-10
C18	-11	C47	-10
C23	-11	C48	-10
C24	-10	C56	-12
C25	-12	C57	-12
C26	-12	C58	-12
C27	-12	C67	-10
C28	-12	C68	-10
C34	-11	C78	-11

- Case 2: Water - 77 kg/min, Air – 0.07 kg/min

As seen in Tab. 8, the domain cross-correlation lags for the capacitances are at around -13, ignoring the "0, -1,-2,-3" lags. 23 lags are corresponding to approximately 100 to 150ms time. This yields a calculated flow velocity of 0.187/0.13 = 1.43 m/s.

Table 8: Cross-correlation for Water - 77 kg/min, Air – 0.07 kg/min

Normalized Capacitance	Peak Correlation Lag	Normalized Capacitance	Peak Correlation Lag
C12	-15	C35	-15
C13	-15	C36	-15
C14	-12	C37	-15
C15	-15	C38	-15
C16	-1	C45	-13
C17	-2	C46	-13
C18	-3	C47	-12
C23	-14	C48	-12
C24	-12	C56	-14
C25	-14	C57	-2
C26	-14	C58	0
C27	-14	C67	-10
C28	-14	C68	-10
C34	-13	C78	-11

6.2.3. Slug

- Case 1: Water - 75 kg/min, Air – 0.3 kg/min

Tab. 9 suggests the domain cross-correlation lags for the capacitances are at around -6. 28 lags are corresponding to approximately 50 to 70ms time. Therefore, the flow velocity is estimated to be 0.187/0.06 = 3.11 m/s for this experiment.

Table 9: Lags based on peaks of cross-correlation for Water - 75 kg/min, Air – 0.3 kg/min

Normalized Capacitance	Peak Correlation Lag	Normalized Capacitance	Peak Correlation Lag
C12	-6	C35	-6
C13	-6	C36	-6
C14	-6	C37	-6
C15	-6	C38	-6
C16	-6	C45	-6
C17	-6	C46	-6
C18	-6	C47	-6
C23	-5	C48	-6
C24	-6	C56	-7
C25	-6	C57	-7
C26	-6	C58	-7
C27	-6	C67	-5
C28	-6	C68	-5
C34	-6	C78	-6

- Case 2: Water - 77 kg/min, Air – 0.5 kg/min

Tab. 11 suggests the domain cross-correlation lags for the capacitances are at around -4. 28 lags are corresponding to approximately 30 to 50ms time. Hence, this experiment's flow velocity is estimated as 0.187/0.04 = 4.67 m/s.

The average flow velocities for three flow regimes are shown in Tab. 10. The average flow velocity of Slug regime is in the expected region. A pattern can be seen with the flow velocity increasing as the flow changes from complex flow regimes of Plug to Slug.

Table 10: Average flow velocity from cross-correlation technique on dual-plane ECT

Flow Regime	Average Flow Velocity (m/s)
Annular	1.43 to 1.87
Plug	1.24 to 1.43
Slug	2.67 to 3.74

6.3 Volume Ratio Estimation

In the volume ratio estimation FNN model, a total of 1350 samples are used for training and testing. These samples are partitioned into training, validation, and testing datasets with a 70:15:15 ratio. The activation function in the hidden layer is *tansig,* and in the output layer is linear. The model's performance, shown in Fig. 14, achieves an R-value of 0.95 for the test dataset, with an overall R-value also standing at 0.95. Some of the outputs are far from the target since the regression model is not perfect and has an R^2 value of 0.9. This model can be used to estimate the volume ratio in the pipe with good confidence.

Table 11: Cross-correlation for Water - 77 kg/min, Air – 0.5 kg/min

Normalized Capacitance	Peak Correlation Lag	Normalized Capacitance	Peak Correlation Lag
C12	-5	C35	-4
C13	-4	C36	-4
C14	-4	C37	-4
C15	-4	C38	-4
C16	-4	C45	-4
C17	-4	C46	-5
C18	-4	C47	-4
C23	-3	C48	-4
C24	-4	C56	-4
C25	-4	C57	-4
C26	-5	C58	-4
C27	-5	C67	-4
C28	-5	C68	-4
C34	-5	C78	-4

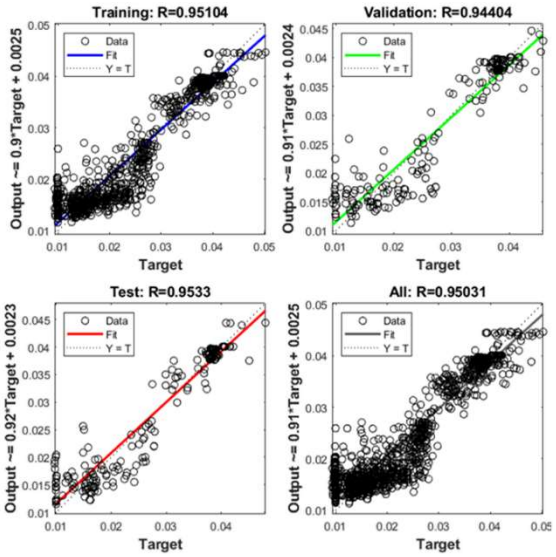


Figure 14: Performance of volume ratio estimation model with PDT120, PDT121 and FT131 as inputs and 1/Volume Ratio from ECT as output

7. Summary and Discussions

The data-driven multiphase flow metering models developed, capable of classifying flow regimes and estimating phase fractions and velocities for two-phase air/water flow, are developed after collecting ECT data from the horizontal flow rig located at USN.

The flow regime classification model, using ECT, achieved an accuracy surpassing 94%. Additionally, a sensor fusion model integrating ECT and pressure sensor data for flow regime classification exceeded 98% accuracy. For annular, plug and slug regimes, flow velocity was estimated using cross-correlation. The volume ratio estimation neural network model attained an R-value greater than 0.95.

This paper demonstrates the feasibility of multiphase flow metering through the use of ECT and pressure sensor data. As depicted in Fig. 15, the data acquired from these sensors can be directly channelled into dedicated ML algorithms to provide insights into multiphase flow in pipeline. This approach facilitates the monitoring and control of processes involving multiphase flow with real time processing of process data on premises or in the enterprise cloud.

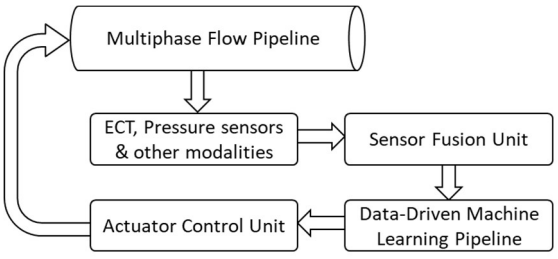


Figure 15: An algorithm for actuator control using a data-driven ML metering model.

Applying ML algorithms to time-series data from these sensors eliminates the necessity for complex mathematical time-series and image-processing methods.

Acknowledgment
Many thanks to Fredrik Hansen of USN for his support in the operation of the multiphase flow rig. This work is closely coupled to an ongoing project SAM (SAM: Self Adapting Model-based system for Process Autonomy - SINTEF), with Equinor and SINTEF as the major partners.

References

Almalki, N., Ahmed, W.H., 2020. Prediction of two-phase flow patterns using machine learning algorithms. https://doi.org/10.11159/ffhmt20.182

Alssayh, M., Addali, A., David, Dao, T., 2013. Identification Of Two Phase Flow Regime Using Acoustic Emission Technology.

Dupré, A., Ricciardi, G., Bourennane, S., Mylvaganam, S., 2017. Electrical Capacitance-Based Flow Regimes Identification—Multiphase Experiments and Sensor Modeling. IEEE Sens. J. 17, 8117–8128. https://doi.org/10.1109/JSEN.2017.2707659

Godfrey Nnabuife, S., Kuang, B., Whidborne, J.F., Rana, Z., 2021. Non-intrusive classification of gas-liquid flow regimes in an S-shaped pipeline riser using a Doppler ultrasonic sensor and deep neural networks. Chem. Eng. J. 403, 126401. https://doi.org/10.1016/j.cej.2020.126401

H. L. M. Ameran, E. J. Mohamad, R. A. Rahim, W. N. A. Rashid, M. M. Mohamad, H. Hashim, Z. Zakaria, M. F. Shaib, O. M. F. Marwah, 2015. Velocity Measurement Simulative Study Of Twin Plane ECT Using Advanced Cross Correlation Technique. ARPN J. Eng. Appl. Sci., 19 10. https://doi.org/October 2015

Hansen, L.S., Pedersen, S., Durdevic, P., 2019. Multi-Phase Flow Metering in Offshore Oil and Gas Transportation Pipelines: Trends and Perspectives. Sensors 19, 2184. https://doi.org/10.3390/s19092184

Jayanti, S., 2011. Wavy Flow. Begel House Inc. https://doi.org/10.1615/AtoZ.w.wavy_flow

Liné, A., Fabre, J., 2011. Stratified Gas-Liquid Flow. Begel House Inc. https://doi.org/10.1615/AtoZ.s.stratified_gas-liquid_flow

Noorain Syed Kazmi, 2023. Multiphase Flow Metering with Multimodal Sensor Suite for Identification of Flow Regimes, and Estimation of Phase Fractions and Velocities. University of South-Eastern Norway, Porsgrunn, Norway.

Pereyra, E., Torres, C., Mohan, R., Gomez, L., Kouba, G., Shoham, O., 2012. A methodology and database to quantify the confidence level of methods for gas–liquid two-phase flow pattern prediction. Chem. Eng. Res. Des. 90, 507–513. https://doi.org/10.1016/j.cherd.2011.08.009

Process Tomography Limited, 2011. Electrical Capacitance Tomography System Type TFLR5000 Operating Manual.

Saied, I., Meribout, M., 2016. Electronic hardware design of electrical capacitance tomography systems. Philos. Trans. R. Soc. Math. Phys. Eng. Sci. 374, 20150331. https://doi.org/10.1098/rsta.2015.0331

Soo, S.L., 1990. Multiphase fluid dynamics. Science Press ; Gower Technical, Beijing : Aldershot ; Brookfield, USA.

Stavland, S.H., Arellano, Y., Hunt, A., Maad, R., Hjertaker, B.T., 2021. Multimodal Two-Phase Flow Measurement Using Dual Plane ECT and GRT. IEEE Trans. Instrum. Meas. 70, 1–12. https://doi.org/10.1109/TIM.2020.3034615

Tan, C., Dong, F., 2023. Sensor Instrumentation for Flow Measurement, in: Narayan, R. (Ed.), Encyclopedia of Sensors and Biosensors (First Edition). Elsevier, Oxford, pp. 536–554. https://doi.org/10.1016/B978-0-12-822548-6.00074-1

Vohr, J.H., 1960. Flow Patterns Of Two-Phase Flow–A Survey Of Literature (No. TID-11514; CU-2-60-AEC-187-Ch.). Columbia Univ., New York. Engineering Research Labs.

Wang, H.X., Zhang, L.F., 2009. Identification of two-phase flow regimes based on support vector machine and electrical capacitance tomography. Meas. Sci. Technol. 20, 114007. https://doi.org/10.1088/0957-0233/20/11/114007

Zeigarnik, Y.A., 2011. Annular flow. Begel House Inc. https://doi.org/10.1615/AtoZ.a.annular_flow

Simulation of Oil Recovery Through Advanced Wells Using a Transient Fully Coupled Well-Reservoir Model

Madhawee Anuththara[1], Ali Moradi[1], Amaranath S. Kumara[1], Britt M. E. Moldestad[1]

[1]*Department of Process, Energy and Environmental Technology, University of South-Eastern Norway, Norway.*
anuththaragr@gmail.com, ali.moradi@usn.no, amaranath.s.kumara@usn.no, britt.moldestad@usn.no

Abstract
Oil recovery can be enhanced by maximizing the well-reservoir contact using long horizontal wells. One of the main challenges of using such wells is the early breakthrough of unwanted fluids due to the heel-toe effect and heterogeneity along the well. To tackle this problem, advanced wells are widely applied today. The successful design of such wells requires an accurate integrated dynamic model of the well and reservoir. This paper aims at developing appropriate integrated well-reservoir models for achieving optimal long-term oil recovery from advanced well models.

In this study, OLGA® which is a dynamic multiphase flow simulator is implicitly coupled to ECLIPSE™ which is a dynamic reservoir simulator for developing accurate models to simulate oil production from advanced wells under various production/injection strategies. A realistic heterogeneous light oil reservoir with an advanced horizontal well is used as a case study. Flow Control Devices (FCDs) are the key component of advanced wells and the functionality of the main types of FCDs in improving the oil production, minimizing the cost and carbon footprint is investigated.

According to the obtained results, by implementation of FCDs the water breakthrough time is delayed by 180 days and the cumulative water production with ICD, AICD, and AICV completions is reduced by 26.8%, 33.1%, and 49.1%, respectively, compared to the open-hole case. Besides, the results show that linking OLGA and ECLIPSE is a numerically stable and accurate approach for modeling the interaction between the dynamic reservoir and dynamic well behavior for simulation oil recovery from advanced wells.

Keywords: Advanced well, ICD, AICD, AICV, OLGA-ECLIPSE coupling

1. Introduction

The DNV Energy Transition Outlook 2022 projects that oil, and gas will still fulfill 39% of the world's energy needs in 2050 (DNV, 2022). Therefore, in an energy transition period, improving the efficiency of the oil recovery methods is important for several reasons. The improved efficiency of the oil recovery methods can lead to cost savings. Moreover, enhancing the oil recovery methods is important to maximize the amount of oil that can be extracted from existing fields so that the resources can be utilized as efficiently as possible (Aakre *et al*., 2013).

To maximize the oil production and recovery, it is important to obtain maximum reservoir contact and to prevent the negative effects of early gas or water breakthroughs. Long horizontal wells can be used to achieve this goal (Aakre *et al*., 2013).. However, there are some challenges associated with horizontal wells, such as early gas/water breakthrough, caused by the water coning effect towards the heel due to the heel-toe effect and heterogeneity along the horizontal well (Moradi *et al*., 2020). To address this issue, inflow control technologies like passive inflow control devices (ICDs), autonomous inflow control devices (AICDs), and autonomous inflow control valves (AICVs) are widely used in oil well completion (Birchenko *et al*., 2010; Aakre *et al*., 2013).

ICDs can balance the drawdown pressure along the horizontal well, thus preventing an early water breakthrough, but they cannot choke the water once it eventually enters the well. The use of AICDs will provide both a delay in the early water breakthrough as well as the possibility of partially choking back water or gas automatically after the breakthrough. AICVs are designed to delay the early breakthrough behaving like AICD until the breakthrough and they can almost completely choke back water or gas autonomously after the breakthrough. Consequently, applying inflow control technologies in horizontal well completions and using Enhanced Oil Recovery (EOR)/Improved Oil Recovery (IOR) technologies would have significant potential to extract non-recoverable oil resources cost-effectively (Mathiesen et al., 2011; Moradi *et al*., 2020; Moradi *et al*., 2022; Moradi, Moldestad and Kumara, 2023).

Before implementing new technologies in an existing reservoir, conducting oil production simulations is standard practice. OLGA is a dynamic multiphase flow simulator for production wells and ROCX and ECLIPSE are reservoir simulation tools. By coupling OLGA with ROCX or ECLIPSE, multi-phase flow behavior in the total oil production can be simulated (Moradi *et al*., 2022; Moradi, Moldestad and Kumara, 2023). The ROCX software is unable to simulate reservoirs with IOR methods such as water flooding while ECLIPSE does offer this capability. ROCX simulations also tend to have relatively longer computation times compared to ECLIPSE. Moreover, ROCX can be used to model near-wellbore reservoir, but ECLIPSE has the facility to model the full reservoir (Schlumberger, 2020). Many studies have focused on linking ROCX to OLGA due to the limited specifications required. However, there is a research gap when it comes to the coupling of ECLIPSE and OLGA for simulation of oil production through advanced wells. This paper aims to provide more insight into the simulation of oil recovery from advanced wells by developing transient fully coupled well-reservoir models using OLGA and ECLIPSE.

2. Inflow control technologies

Horizontal wells often face issues like water and gas coning, as well as early water breakthroughs due to reservoir heterogeneity and the heel-toe effect. To address these challenges, passive and autonomous inflow control technologies have been introduced. By implementing these technologies in horizontal wells, balanced drainage can be achieved, leading to increased oil production and improved recovery rates.

2.1. Passive inflow control devices (ICD)

ICD limiting the flow by creating an additional pressure drop to achieve an evenly distributed flow profile along a horizontal well as shown in Fig. 1. This pressure drop is a function of the liquid flow rate, the density of the fluid, and the viscosity of the fluid, though the viscosity plays a less important role.

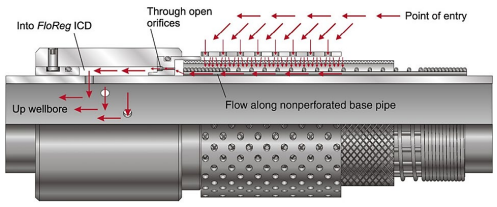


Figure 1: Orifice(nozzle) type ICD (Birchenko, Muradov and Davies, 2010).

As a result of an even production rate along the well, water/gas breakthrough could be delayed significantly. Specifically, ICDs are designed to apply a specific differential pressure at a specified flow rate through the device. The main disadvantage of passive ICDs is that they cannot choke back the water after the breakthrough. In this situation, the whole well is choked in order to prevent the increase of the water cut, greater than the capacity of the separation facilities, which in turn results in a reduction in oil production (Moradi and Moldestad, 2020). This study uses the orifice (nozzle) type ICDs. The orifice type ICDs create a resistance when the fluid tries to enter the well, by forcing the flow through a set of small-diameter nozzles or orifices. The governing equation of the nozzle-type ICD, derived by Bernoulli's equation, is as follows (Moradi and Moldestad, 2020):

$$\dot{Q} = C_D A \sqrt{\frac{1}{1-\beta^4}} \cdot \sqrt{\frac{2\Delta P}{\rho}} \qquad (1)$$

where $\dot{Q}$ is the volume flow rate of the fluid passing through the ICD, ΔP is the pressure drop over the ICD and, ρ is the fluid density and $\beta = d/D$ (where d and D are the diameters of the orifice and production tubing respectively). C_D is the *discharge coefficient* and it is calculated as; $C_D = A_{vc}/A$. Here, A_{vc} is the minimum jet area just downstream of the orifice called *Vena Contracta*.

2.2. Autonomous inflow control devices (AICD)

To address the limitations of ICDs, that cannot control the water and gas production after breakthrough, AICDs were developed. The AICDs can function as an ICD until a breakthrough occurs, and then automatically control and reduce the water and gas production. The AICD combines passive inflow control with an active control element to produce a pressure drop to autonomously restrict the flow of the unwanted fluid with no need for surface control.

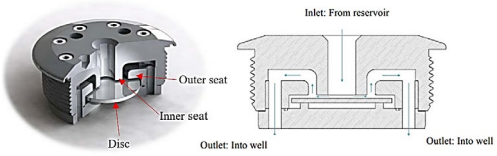


Figure 2: Schematic diagram of Statoil's RCP valve (Mathiesen, Aakre and Werswick, 2011).

Among the various designs of AICDs, the most widely used type is known as Rate Control Production (RCP), developed by Statoil and this study used RCP valves as the AICDs. As shown in Fig. 2, the RCP valve consists of 3 parts, a free-floating disc, an inner seat, and an outer seat. When the valve is in operation, the force acting on the disc is the sum of the pressure forces acting on both sides of the disc. The working method is based on Bernoulli's principle. When more viscous fluids flow through a valve, friction loss increases and the pressure recovery of the dynamic pressure

decreases. As a result, the pressure on the outlet side of the valve (top side of the disc in Fig. 2), decreases, leading to a reduced force on the disc towards the inlet. This causes the disc to move away from the inlet, thereby increasing the flow area available, and boosting the flow rate of the high viscous fluid. This works vice versa for low viscous fluids like water and gas, resulting in autonomously reduced production of unwanted fluids. Statoil developed a governing equation for the differential pressure across the RCP valve, δP and it validated with experimental data, which is:

$$\delta P = f(\rho, \mu) \cdot a_{AICD} \cdot q^{x} \quad (2)$$

Where, a_{AICD} and x are user input model constants, which depend on different RCP designs for different oil fields and their fluid properties. The function $f(\rho, \mu)$ is an analytic function of the fluid mixture density ρ and viscosity μ, defined as:

$$f(\rho, \mu) = \left(\frac{\rho_{mix}^2}{\rho_{cal}}\right) \cdot \left(\frac{\mu_{cal}}{\mu_{mix}}\right)^{y} \quad (3)$$

Here, y is a user-defined constant, ρ_{cal} and ρ_{mix} are calibration and mixture density and μ_{cal} and μ_{mix} are calibration and mixture viscosity, and they can be defined as follows, while α is the volume fraction of each phase:

$$\rho_{mix} = \alpha_{oil}\,\rho_{oil} + \alpha_{water}\,\rho_{water} + \alpha_{gas}\,\rho_{gas} \quad (4)$$

$$\mu_{mix} = \alpha_{oil}\,\mu_{oil} + \alpha_{water}\,\mu_{water} + \alpha_{gas}\,\mu_{gas} \quad (5)$$

2.3. Autonomous inflow control valves (AICV)

AICV is a new type of inflow control device developed by InflowControl AS, and it can equalize the inflow before the breakthrough like AICD. As opposed to AICDs, which can partially close against unwanted fluids, AICVs can almost completely choke low-viscosity fluid, such as water or gas.

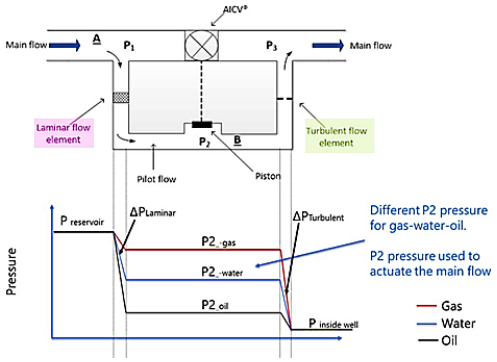


Figure 3: Simplified sketch of the flow paths in AICV and pressure changes inside for different fluids (Aakre, Mathiesen and Moldestad, 2018).

AICVs are fully self-regulating and do not rely on any external control systems and are designed to achieve the autonomous functionality by distinguishing between fluids based on their density and viscosity. The fundamental theory behind the AICV operation is the difference between the pressure drop in a laminar flow restrictor and a turbulent flow restrictor shown in Fig. 3. The laminar flow restrictor is like a pipe segment, and pressure drop across a laminar flow restrictor $\Delta P_{Laminar}$ can be expressed, as a relation of fluid viscosity μ, velocity v, pipe length L and pipe diameter D (Aakre *et al.*, 2013).

$$\Delta P_{Laminar} = \frac{32 \cdot \mu \cdot \rho \cdot v \cdot L}{D^2} \quad (6)$$

The turbulent flow restrictor can be considered as an orifice plate, and the pressure drop across the turbulent flow restrictor $\Delta P_{Turbulent}$ can be expressed as a relation of fluid density ρ, velocity v, and geometric contact K (Aakre *et al.*, 2013).

$$\Delta P_{Turbulent} = K \cdot \frac{1}{2} \cdot \rho \cdot v^2 \quad (7)$$

According to these relationships, $\Delta P_{Laminar}$ depends on the viscosity the fluid, while $\Delta P_{Turbulent}$ depends on the density of the fluid. When a viscous fluid such as oil passes through a laminar flow restrictor, it experiences a greater pressure drop than fluids with a low viscosity such as water and gas. A low-viscosity fluid, on the other hand, experiences a lesser pressure drop across the laminar flow restrictor, resulting in a higher pressure in chamber 'B' (P2) in Fig. 3. Due to the high pressure, a piston in chamber 'B' will be actuated, closing the valve. AICVs are designed based on these principles to remain fully open for oil while almost completely closed to prevent the flow of unwanted fluids.

3. Multi-segment well model (MSW)

The Multi-Segment Well model is a special extension available in ECLIPSE that offers comprehensive and accurate modeling facilities for the fluid behavior in advanced wells. There is a complex relationship between pressure gradients and changes in fluid composition induced by specific components of advanced wells. The MSW can be used to model this behavior. This model divides the production tubing into several one-dimensional segments. There is a node and a flow path, and each segment contains its own set of independent variables to describe the fluid conditions in that region. The variables for each segment are evaluated by solving material balance equations for each phase or component, and using the pressure drop equation that incorporates local hydrostatic, frictional, and acceleration pressure gradients (Schlumberger, 2020; Moradi *et al.*, 2022; Moradi, Moldestad and Kumara, 2023).

4. Development of the OLGA/ECLIPSE model

OLGA serves as a dynamic multiphase flow simulator for the production well, while ECLIPSE functions as a reservoir simulator that can be integrated with OLGA as a plug-in. The combination of OLGA and ECLIPSE provides a tool for modeling and simulating multiphase flow from the reservoir pore to the production pipeline.

4. 1. Development of the reservoir model in ECLIPSE

4.1.1. Grid

The dimensions of the synthetically designed reservoir using MRST are mentioned in Tab. 1.

Table 1: Dimensions of the reservoir.

Dimension	Value
Length of the reservoir (x)	1500m
Width of the reservoir (y)	500m
Height of the reservoir (z)	50m

The horizontal oil production well is positioned in the x-direction of the reservoir (length), 5 m below the top of the reservoir. For improved oil recovery, a horizontal water injection well with 20 perforations is used and it is positioned in the x-direction, 45 m below the top of the reservoir.

Table 2: Number of cells and their sizes in the grid.

Direction	Number of cells	Size of the cells
x	nx = 30	50 m (constant)
y	ny = 10	50 m (constant)
z	nz = 5	10 m (constant)

Generally, FCDs are installed with a sand screen and the length of one production joint is 12.4 m of the well. Since the reservoir length (x-direction) is 1500 m, 120 FCDs can be placed along the well. However, it is complex to simulate the real well with a huge number of FCDs as it consumes a long simulation time. Therefore, one equivalent FCD is used to represent 4 real FCDs. Thus, 30 cells are considered in x-direction and 30 FCDs are used along the well. In y and z-directions, 10 and 5 cells are considered respectively. The grid settings in ECLIPSE, including the number of cells in each direction and their sizes are given in Tab. 2. The 3D view of the reservoir and wells completed with FCDs is given in Fig. 4.

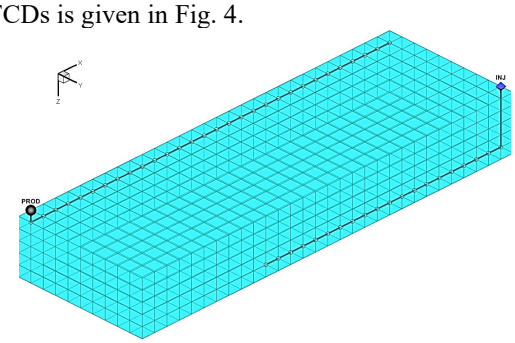


Figure 4: 3D view of the reservoir with wells.

4.1.2. The fluid and rock properties of the reservoir

It is assumed that the synthetically designed reservoir has conditions similar to the Troll field in the North Sea, containing a viscous oil with a viscosity of 2.7 cP. Therefore, the reservoir fluid can be considered as black oil type (oil viscosity is 2 to 3 – 100 and up). Reservoir fluid properties and some rock properties used for the OLGA/ECLIPSE model are listed in Tab. 3.

Table 3: Fluid and rock properties of the reservoir.

Property	Value
Oil density	950 kg/m^3
Oil viscosity	2.7 cP
Water density	1100 kg/m^3
Gas density	0.67 kg/m^3
Solution GOR	50 Sm^3/Sm^3
Porosity	0.15-0.27
Initial water saturation	0.12
Reservoir pressure	130 bara
Reservoir temperature	68 ^{0}C

4.1.3. Relative permeability

The reservoir is considered as a heterogeneous sandstone reservoir. In this study, the *log-normal absolute permeability* of the reservoir is assumed in the range 100 - 800 mD s to account for the uncertainty in the reservoir.

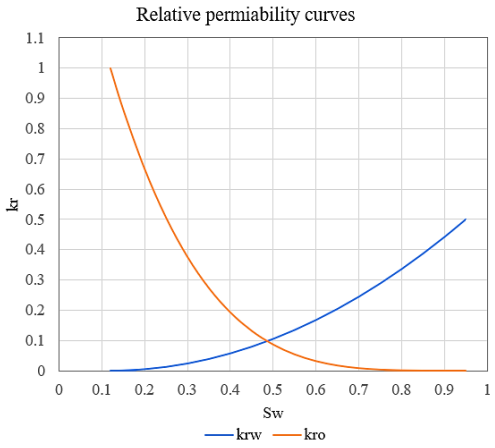


Figure 5: Generated relative permeability values.

The generalized Corey model can be used to calculate the *relative permeabilities* of oil and water using the ECLIPSE software, and the generated relative permeability values are plotted in Fig. 5 where, krw and kro are the relative permeabilities of water and oil respectively.

4.1.3. Initial and boundary conditions

The reservoir model in ECLIPSE assumes an initial oil saturation of 0.88, water saturation of 0.12, and no gas saturation. The production well is regulated with a constant Bottom Hole Pressure (BHP) of 115 bar. With a mean porosity of 0.21, the total void volume of the reservoir is calculated as 7875000 m^3. For 1500 days in operation, approximately two-thirds of the reservoir liquid is expected to be produced. Therefore, the required water injection flow rate by a single injection well is estimated to be 3500 m^3/day. However, this flow rate cannot be

applied due to the industry's maximum allowable injection pressure limitation of 180 bar. Therefore, it was decided to inject water through 20 similar perforations in the horizontal water injection well, each one with a water flow rate of 175 m^3/day. Furthermore, in practical oil and gas production, the total liquid production from a well can be limited by the maximum capacity of the surface facilities. In the study, for the open-hole case model, the maximum liquid production rate is set to 2400 m^3/day.

4.2. Development of well model in OLGA

In the OLGA model, production well consists of two parts: wellbore, and production tubing. It is specified as both pipes are made with the same material combination, where the internal pipe is made of 9 mm thickness of API 5L Grade B carbon steel and other layers consist of two 2 cm concrete layers as shown in Fig. 6.

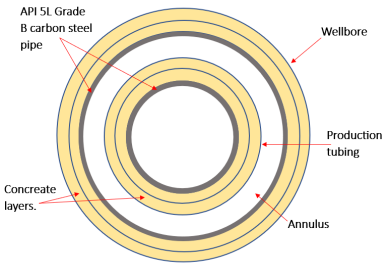


Figure 6: Material structure of wellbore and production tubing.

4.2.1. Tables and curves

By performing non-linear curve fitting for experimental data, the relationship of the autonomous functions of AICD/AICV with respect to the Water Cut (WC) can be determined (Moradi *et al.*, 2022; Moradi, Tavakolifaradonbe and Moldestad, 2022).

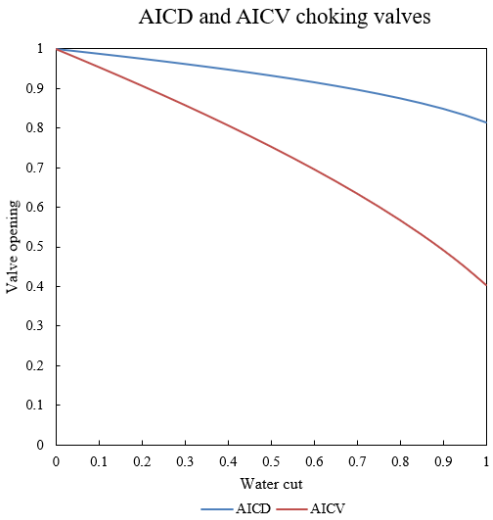


Figure 7: AICD and AICV choking valves for oil viscosity 2.7 cP for 15 bar pressure drop.

These autonomous functions of FCDs are implemented in the OLGA models, based on the pressure drawdown, 15 bar in this case, by employing a table controller and a transmitter for each FCD.

This table controller gets the measured WC data from the transmitter and provides corresponding control signals to partially close the FCDs for choking the fluid passing through them (Moradi *et al.*, 2022). The generated valve opening values of AICD and AICV with respect to WC is plotted in Fig. 7.

4.2.2. Flow component

Fig. 8 shows the simplified sketch of one oil production zone in OLGA model.

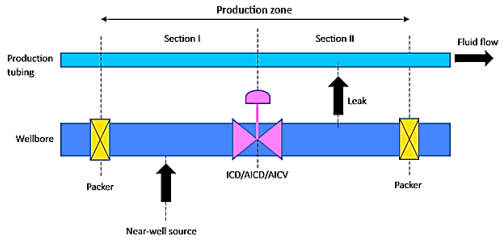


Figure 8: Simplified sketch for one oil production zone.

The production zones are separated by packers to prevent reservoir fluid from flowing in between adjacent zones through annulus. The near-well source in the OLGA model is used to connect OLGA with ECLIPSE accordingly. Then the fluid enters the wellbore through section I after passing through the FCD in Fig. 10. The fluid that enters the wellbore passes to the production tubing via the leak in section II. This setup was proposed by Haarvard Aakre in 2012 and this method has been used for many research (Moradi and Moldestad, 2020).

To develop the OLGA model, two flow paths are required for the wellbore and the production piping with a length of 1500 m for each. As the internals of wellbore and production tubing are made out of API 5L Grade B carbon steel, absolute roughness is considered as 4.572 x 10^{-5} m for both pipes (NEELCONSTEEL, 2022). The diameter of the production tubing and wellbore are assumed as 0.1397m and 0.2159m, respectively. It is assumed that oil is produced from 30 zones in the well, each of which contains two hypothetical sections as shown in Fig. 10. The production well has 30 FCDs. Since one valve is equivalent to 4 real valves, the diameter of one valve (ICD/AICD/AICV) is 0.0042 m considering the Discharge Coefficient (CD) as 0.85. When the valves are not implemented in the horizontal well, it is called “open-hole” completion. which is in a fully open state. The open-hole diameter is set as 0.12 m considering CD as 0.85. Under the case conditions, it is set to run the model

for 1500 days with a minimum time step of 0.00001 seconds and a maximum time step of 1000 seconds. To solve the mass equations, a first-order discretization scheme is selected.

4.2.3. Boundary conditions

Boundary conditions of the OLGA model are set as following Tab. 5.

Table 5: Boundary conditions of the OLGA model.

Flow path name	Boundary Name	Boundary Type in OLGA
Wellbore	Inlet	Closed node
	Outlet	Closed node
Production tubing	Inlet	Closed node
	Outlet	Pressure node, Pressure =115 bar, Temp. = 68˚C

4. Results and discussion

4.1. Results validation with multi-segment well (MSW) model

Since the OLGA-ECLIPSE combination is a new approach, a result validation can be performed to prove its accuracy compared to other modeling and simulation methods. A case was considered for the oil recovery from an advanced horizontal well with AICD well completion followed by vertical water flooding. As shown in Fig. 9, the results obtained by the MSW model and the linked OLGA-ECLIPSE model, are overlapping and this implies that the effort on coupling OLGA-ECLIPSE has been successful.

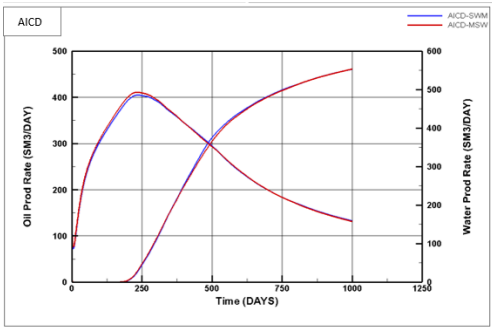


Figure 9: Results validation with MSW model.

4.2. Oil production over water breakthrough

When the oil is produced from a horizontal well, the phenomenon of water coning causes a decrease in the production efficiency. Over time, this leads to an early water breakthrough and a significant reduction in oil production. Typically, the overall oil production gradually increases until a breakthrough occurs. However, once the breakthrough happens, more and more water is pushed toward the well, which in turn suppresses and reduces the oil production. Separating water from the oil during production involves specialized equipment and processes, leading to increased costs. Additionally, the disposal of produced water poses challenges as it often requires treatment to meet environmental regulations. Therefore, delaying water breakthroughs and minimizing water production are crucial to achieve optimal production efficiency and cost reduction in the oil extraction process.

Fig. 10 shows the observed results for the WC over time for different well completions. The open-hole breakthrough occurs on the 620th day of operation while it is on the 800th day for all the other advanced well completions.

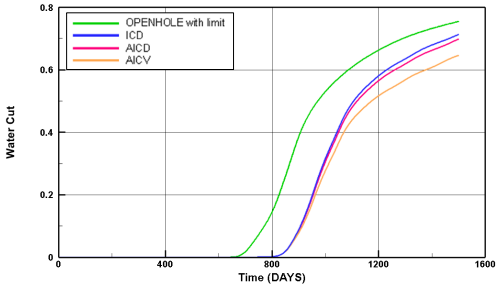


Figure 10: Water cut over the time for different FCD completions.

The implementation of FCDs has significantly delayed the water breakthrough and significantly reduced the total water production over time as expected. This is beneficial for oil recovery with a minimum cost. Until the breakthrough both AICVs and AICDs have behaved like ICDs. After the breakthrough, their autonomous function choked the water considerably and AICVs show their ability to choke more water compared to AICDs.

4.3. Accumulated oil and water production

The simulation results for the accumulated oil and water production are given in Fig. 11. According to the results, compared to the open-hole case, the cumulative oil productions from ICD, AICD, and AICV completions have increased by 2.22%, 1.7%, and 0.2%, respectively, at the end of 1500 days of operation. Moreover, the cumulative water production of ICD, AICD, and AICV is considerably reduced by 26.8%, 33.1%, and 49.1%, respectively, compared to the open-hole case. This indicates that implementation of FCDs in horizontal wells has enhanced the oil recovery to some extent, in addition to the reduction of water production.

Interestingly, the AICV completion has reduced water production by almost half (49.1%), due to the ability of completely choking of low viscous fluids. According to Fig.11, at the end of 1500 days of operation, the WC for AICD and AICV are 0.65 and 0.7 respectively. At this time, based on the vale opening plot in Fig. 9, the value openings for AICD and AICV are 0.95 and 0.65 respectively. This implies that, when the WCs increase with time, the

more choking effects of AICDs and AICVs can be expected.

Moreover, it can be noted that, according to the cumulative oil production, the open-hole case initially has a higher oil production compared to the other well completions. But, due to the early water breakthrough after 620 days (open-hole case), the water that enters the wellbore has suppressed the oil production, resulting in higher accumulated water production and lesser accumulated oil production at the end of the operation.

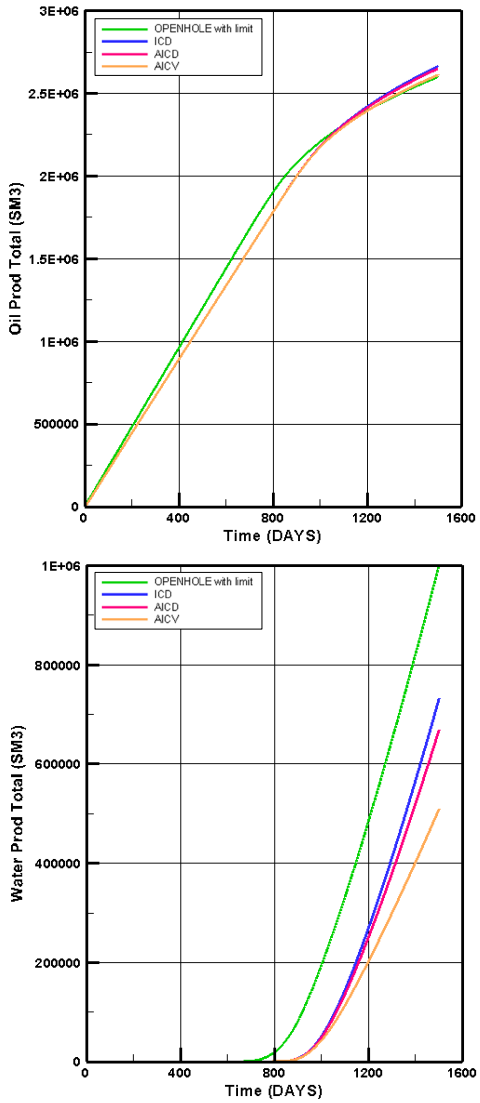


Figure 11: Accumulated oil and water production for open-hole and for different FCD completions.

4.4. Oil and water production rate

The simulation results observed for oil and water production rates are given in Fig. 12. Considering the oil production rates, the open-hole completion initially has the maximum oil production rate (~2265 Sm^3/d) compared to other advanced wells, and that production rate lasts until the water breakthrough only. Although the other advanced well completions have a 5.61% lower oil production rate at the beginning, it lasts for a longer period since the water breakthrough is delayed in advanced wells. But at the end of 1500 days of operation, the OPENHOLE case has achieved the lowest oil production rate as it does not have control over the water production after the breakthrough. And the open-hole case has also the highest water production rate from the beginning. It is generally undesirable to have a high total liquid flow rate. This is because there is then a need for larger surface production facilities to handle the increased liquid volume and higher costs associated with water separation. Ultimately, this situation leads to reduced revenue.

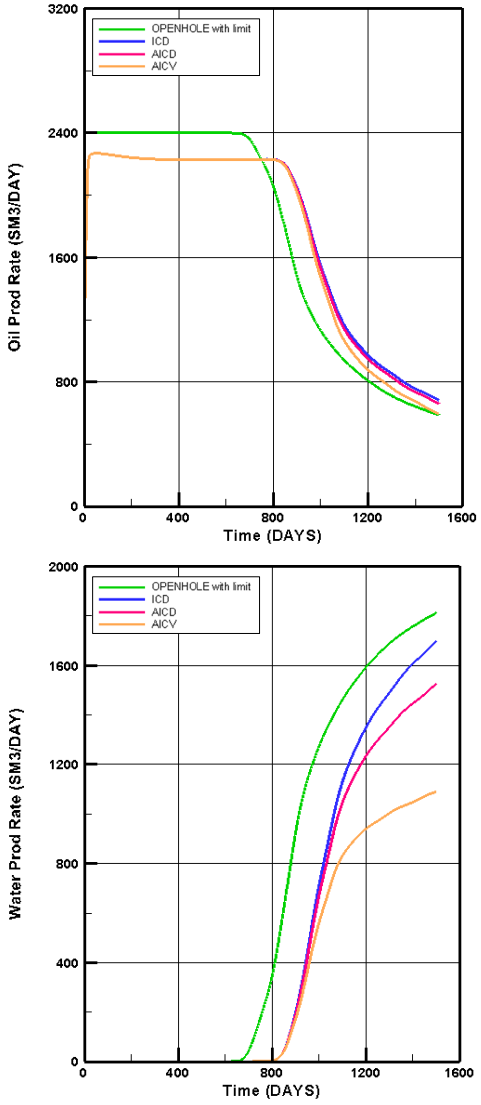


Figure 12: Volumetric oil production rates for open-hole and for different FCD completions.

Considering volumetric flow rates at the end of 1500 days, the ICD, AICD, and AICV completions have achieved 16%, 12.1%, and 1.3% increments in the oil production compared to the open-hole case. But end water production rates of ICD, AICD, and AICV

completions have reduced by 6.2%, 15%, and 39.8% compared to the open-hole case. It appears that despite advanced well completions having a small impact on oil production rates because of the total liquid production limit, advanced wells can significantly reduce the water production by improving the oil production process in a cost-effective manner.

5. Conclusion

To achieve cost-effective oil production, it is important to address the problem of early water breakthrough in horizontal wells. Implementation of ICDs, AICDs, and AICVs evens out the inflows along the well and delays the water breakthrough. The AICDs and AICVs show similar behavior to ICDs before the breakthrough. Advanced wells equipped with FCD completions result in a significant decrease in the production of water after the breakthrough and with a little increase of accumulative oil production compared to the open-hole completion, while AICVs show the best performance in choking water.

The autonomous function of AICD and AICV can be clearly seen if the WC exceeds around 0.9. But the oil production should last longer than 1500 days in order to achieve a higher WC. Therefore, it is recommended to extend the simulation period to observe the true impact of utilizing advanced well technologies for achieving more efficient oil production processes. Realistic results obtained from the simulations in this study indicate that the coupling of the well simulator OLGA and reservoir simulator has been a successful effort in simulating total oil production and this combination can be further applied to more advanced scenarios to compare its effectiveness with other oil production simulators.

Acknowledgment

We gratefully acknowledge the economic support from the Research Council of Norway and Equinor through Research Council Project No. 308817, ‘‘Digital Wells for Optimal Production and Drainage’’ (DigiWell), and for the university of South-Eastern Norway for providing the necessary software arrangements for this work.

References

Aakre, H. *et al.* (2013) ‘Smart Well with Autonomous Inflow Control Valve Technology’, in. *SPE Middle East Oil and Gas Show and Conference*, OnePetro, doi: 10.2118/164348-MS.

Aakre, H., Mathiesen, V. and Moldestad, B. (2018) ‘Performance of CO2 flooding in a heterogeneous oil reservoir using autonomous inflow control’, *Journal of Petroleum Science and Engineering*, 167, pp. 654–663. doi: 10.1016/j.petrol.2018.04.008.

Birchenko, V.M., Muradov, K.M. and Davies, D.R. (2010) ‘Reduction of the horizontal well’s heel-toe effect with inflow control devices’, *Journal of Petroleum Science and Engineering*, 75(1–2), pp. 244–250. doi: 10.1016/j.petrol.2010.11.013.

DNV (2022) *DNV Energy Transition Outlook 2022: A Global and Regional Forecast to 2050*.

Mathiesen, V., Aakre, H. and Werswick, B. (2011) ‘The Autonomous RCP Valve-New Technology for Inflow Control In Horizontal Wells’, in *The Autonomous RCP Valve-New Technology for Inflow Control In Horizontal Wells*. *The SPE Offshore Europe Oil and Gas Conference and Exhibition*, Aberdeen, UK: OnePetro. doi: 10.2118/145737-MS.

Moradi, A. *et al.* (2022) ‘Evaluating the performance of advanced wells in heavy oil reservoirs under uncertainty in permeability parameters’, *Energy Reports*, 8, pp. 8605–8617. doi: 10.1016/j.egyr.2022.06.077.

Moradi, A. and Moldestad, B. (2020) ‘Near-well simulation of oil production from a horizontal well with ICD and AICD completions in the Johan Sverdrup field using OLGA/ROCX’, in *The 61st SIMS Conference on Simulation and Modelling SIMS 2020, Virtual Conference*. Finland, pp. 249–256. doi: 10.3384/ecp20176249.

Moradi, A., Moldestad, B.M.E. and Kumara, A.S. (2023) ‘Simulation of Waterflooding Oil Recovery With Advanced Multilateral Wells Under Uncertainty by Using MRST’, in. *SPE Reservoir Characterisation and Simulation Conference and Exhibition*, OnePetro. doi: 10.2118/212700-MS.

Moradi, A., Tavakolifaradonbe, J. and Moldestad, B.M.E. (2022) ‘Data-Driven Proxy Models for Improving Advanced Well Completion Design under Uncertainty’, *Energies*, 15(20), p. 7484. doi: 10.3390/en15207484.

NEELCONSTEEL (2022) *API 5L Grade B Pipe*, *NEELCON STEEL INDUSTRIES AN ISO 9001:2015 certified*. Available at: https://www.neelconsteel.com/api-5l-grb-carbon-steel-pipes.html#roughness (Accessed: 3 May 2023).

Norskpetroleum (2022) *Exports of Norwegian oil and gas*, *Norwegianpetroleum.no*. Available at: https://www.norskpetroleum.no/en/production-and-exports/exports-of-oil-and-gas/ (Accessed: 22 February 2023).

NPD (2022) *RESOURCE REPORT 2022, Chapter 2–Remaining petroleum resources*, *Norwegian Petroleum Directorate*. Available at: https://www.npd.no/en/facts/publications/reports/resource-report/resource-report-2022/2-remaining-petroleum-resources/ (Accessed: 22 February 2023).

Schlumberger (2020) *ECLIPSE Technical Description*.

Visualization of Industrial Production Processes using 3D Simulation Software for Enhanced Decision-Making

Akshay Goyal

Mälardalen University and Mälardalen Industrial Technology Centre,
akshay.goyal@mdu.se

Abstract

This paper explores the use of 3D simulation software for visualizing industrial production processes and its potential to enhance decision-making for improved production efficiency, quality, and profitability. Industrial production processes are complex and involve many variables and factors that can interact in unpredictable ways. Visualization helps to simplify these complex interactions, identify patterns and relationships, and enable more informed decision-making.

The research question that guides this paper is: *How can the use of 3D simulation software for visualization of industrial production processes improve decision-making and optimize production efficiency, quality, and profitability?* This paper will investigate the benefits and challenges of using 3D simulation software for visualizing industrial production processes, including the ability to identify bottlenecks, and optimize the production process. Further, the paper examines the role of visualization in enabling more informed decision-making, including the ability to analyze production data and make data-driven decisions. To illustrate this, an industrial automation case study consisting of a manufacturing industry modelled in a 3D simulation software has been presented.

The results of this 3D-simulation model provide insights into the advantages and disadvantages of utilizing 3D simulation software to visualize industrial manufacturing processes. The article further presents the significance of these findings for production managers, engineers, and decision-makers. Thus, the purpose of this study is to help readers understand how using 3D simulation software for visualization of industrial production processes can improve decision-making and optimize production efficiency, quality, and profitability.

Keywords: Industrial production processes, 3D simulation software, visualization, decision-making, production efficiency, industrial automation, data-driven decisions, manufacturing industry, optimization.

1. Introduction

Industrial production processes are central to the production of goods in the manufacturing industry. A series of activities involved in converting raw materials into finished goods using various equipment, tools, and machines are referred to as industrial production processes. Material handling, assembly, packaging, and quality control are examples of these processes. Optimizing industrial production processes is critical to achieving better production efficiency, quality, and profitability. This is because inefficient and suboptimal production processes can lead to bottlenecks, delays, and errors, which can result in increased costs, decreased quality, and reduced competitiveness (Xu et al., 2021). To overcome these challenges, manufacturers are increasingly turning to technology to improve their production processes. 3D simulation software is one such technology that has been gaining popularity in recent years, as it enables manufacturers to visualize and analyze their production processes in a virtual environment before implementing changes in the physical world (Abidi et al., 2020).

Visualization is the use of graphical or pictorial representations to convey complex information in an intuitive and easily understandable way. In the context of industrial production processes, visualization help to simplify complex processes, make patterns and relationships more evident, and enable more informed decision-making (Wang et al., 2015). In the context of industrial production processes, visualization helps to provide a clear understanding of the process flow, identify bottlenecks, simulate, and test different scenarios, and optimize resource allocation. Additionally, visualization can help decision-makers to understand the impact of different decisions on production processes and outcomes and choose the best course of action accordingly (Molenda et al., 2019).

In industrial production processes, there are often a large number of variables and factors to consider, such as machine parameters, production rates, and quality metrics (Xu et al., 2021). These factors can interact in complex ways, making it difficult to understand and optimize the production process (Molenda et al., 2019). Visualization can help to simplify these complex interactions by presenting the information in a way that is easier to understand

and analyze. Furthermore, visualization is used to identify patterns and relationships that may not be apparent from raw data. For example, by plotting machine performance over time, it may be possible to identify recurring issues that are impacting machine production efficiency (Atmakuri et al., 1993). By presenting this information graphically, it is easier to identify patterns and make informed decisions about how to address the issues.

Through displaying information in a way that is clear and accessible, visualization also aids in enabling more informed decision-making (Atmakuri et al., 1993). This can assist stakeholders in making more informed decisions by helping them understand the effects of various actions on production efficiency, quality, and profitability (Yang et al., 2021). Therefore, the topic of visualization of industrial production processes using 3D simulation software for enhanced decision-making is important because it enables manufacturers to optimize their production processes, leading to better production efficiency, quality, and profitability. By visualizing production processes in 3D simulation software, manufacturers can identify bottlenecks, test different scenarios, and make data-driven decisions before implementing changes in the physical world. This can lead to reduced costs, improved quality, and increased competitiveness, all of which are critical factors for success in the manufacturing industry. Hence, the research question has been formulated as follows:

How can the use of 3D simulation software for visualization of industrial production processes improve decision-making and optimize production efficiency?

2. Methodology

The problem addressed in this study is the need for enhanced decision-making in industrial production processes. Traditional methods of decision-making may lack comprehensive insights into the complex dynamics of production systems. Therefore, there is a growing interest in utilizing 3D simulation software to visualize industrial processes and improve decision-making capabilities.

The purpose of this study is to explore the potential benefits of using 3D simulation software for visualizing industrial production processes and examine its impact on decision-making. By investigating this area, the study aims to provide valuable insights and practical recommendations for improving decision-making in industrial settings. Figure 1 illustrates the research methodology involved in the development of the simulation model. A comprehensive review of scholarly and research articles was conducted to gain insights into visualization techniques, simulation software, and decision-making methodologies. The review included a range of sources including scholarly articles, research papers, and industry reports to understand the current state of research in the field. Data collection for this study involved gathering relevant information from various sources. Keywords related to industrial production processes, visualization, and decision-making were used to identify peer-reviewed articles that contributed to the research area. These articles were carefully selected and analysed to extract valuable insights.

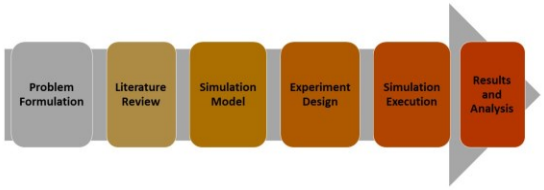


Figure 1: Research Method

The software Visual Components® has been utilized by the author to develop the 3D simulation model. The model represents an industrial production process consisting of injection moulding machines, series of conveyor belts, robots and a warehouse for storage. The model was developed based on the specific requirements and specifications of the case study and a dynamic representation of components flow has been visualized. The experiment design phase involved defining the simulation resources and parameters. Key variables, such as production volumes, cycle times, and system constraints, were determined to create a realistic simulation environment. The experimental design aimed to capture the complexities and dynamics of the industrial production process under study. The simulation model was executed to replicate the production operations. During the simulation run, process states, utilization rates, and performance metrics were recorded. This data provided valuable insights into the functioning of the production system and allowed for a detailed analysis of the simulation results.

The collected data was analysed to answer the research question posed in the study. The analysis focused on evaluating the effectiveness of the visualization of industrial production processes using the 3D simulation software. Performance metrics, such as throughput, cycle times, and resource utilization, were examined to assess the impact of visualization on decision-making. The results were interpreted to draw conclusions and provide practical recommendations. Based on the analysis of the simulation results, conclusions were drawn regarding the effectiveness of using 3D simulation software for visualizing industrial

production processes and enhancing decision-making. The study provides insights into the benefits and limitations of visualization techniques and offers recommendations for further improvements in decision-making practices within industrial settings.

2.1. Simulation of Industrial Production Processes

Industrial production processes are fundamental in the manufacturing industry, encompassing a series of stages and steps that convert raw materials or components into finished products (Molenda et al., 2019). These processes involve crucial operations such as material handling, assembly, packing, and quality control (Chawla & Banerjee, 2001), which have been recognized as essential for simulating manufacturing processes. However, the complexity of these processes, often involving multiple interconnected systems and subsystems, presents challenges in comprehending and optimizing them effectively (Atmakuri et al., 1993).

Optimizing industrial production processes holds immense value as it has the potential to improve various aspects, including production efficiency, product quality, and overall profitability (Yang et al., 2021). By identifying and addressing bottlenecks, minimizing waste, and maximizing throughput, optimization efforts can lead to significant cost savings, enhanced customer satisfaction, and increased competitiveness in the marketplace. However, achieving optimization in industrial production processes is not without its obstacles. Challenges such as demand variability, equipment breakdowns, and workforce availability can impede the optimization journey, necessitating the adoption of advanced technologies to overcome these hurdles (Chawla & Banerjee, 2001; Yang et al., 2021). In this context, the utilization of 3D simulation software and visualization techniques emerges as a powerful solution. These technologies enable a deeper understanding and effective management of industrial production processes by providing insights into complex interactions and facilitating informed decision-making. 3D simulation software allows for the creation of virtual environments where manufacturing procedures can be simulated, analyzed, and optimized (Chawla & Banerjee, 2001). By replicating real-world scenarios, these simulations enable manufacturers to test different scenarios, evaluate the impact of process changes, and identify potential areas for improvement.

Visualization techniques, particularly in 3D, offer a more intuitive and comprehensive representation of the production processes. They surpass the limitations of traditional two-dimensional (2D) images by providing a richer visualization of spatial information and enabling real-time interaction (Molenda et al., 2019). By visualizing the processes in a virtual environment, decision-makers can better grasp the flow of materials, detect inefficiencies or bottlenecks, and make data-driven decisions to optimize the system.

2.2. Building 3D Simulation Model

The use of a 3D manufacturing simulation model provides a valuable tool for simulating and analyzing basic manufacturing procedures within a virtual environment. This technology leverages 3D visualization, which utilizes computer technology to depict real-world objects in a virtual space, surpassing the limitations of two-dimensional (2D) images in representing complex spatial information (Xu et al., 2021). By incorporating real-time interaction capabilities, the 3D visualization enhances the immersive experience and enables users to engage with the simulated environment effectively. The implementation and integration of a 3D simulation model involve several important steps. Key considerations during this process include defining resource parameters, selecting appropriate equipment, establishing fundamental simulation logic, and incorporating relevant production data (Xu et al., 2021). These elements are crucial for creating an accurate and realistic representation of the manufacturing system within the simulation model.

The resource parameters encompass various factors such as material properties, production capacities, and operating constraints. These parameters define the characteristics and capabilities of the resources involved in the manufacturing process, enabling the simulation model to accurately reflect their behavior and interactions (Chawla & Banerjee, 2001). The selection of equipment involves identifying and configuring the machinery, tools, and systems that are integral to the manufacturing process. By modelling these equipment components within the simulation, their functionalities and interactions can be evaluated and optimized (Wang et al., 2015). Fundamental simulation logic refers to the underlying principles and algorithms that govern the behavior and dynamics of the simulated manufacturing process.

Creating a 3D simulation of an industrial production procedure entails multiple processes (Figure 2). First, describe the simulation's scope and objectives, outlining the precise goals to be attained. Next, collect relevant data about the manufacturing process and identify the components and processes that must be modelled. Choose a 3D simulation program that matches the project's criteria. Create the simulation model by combining the acquired data and modelling the manufacturing line's

components and procedures. This includes developing 3D models, setting material flow rules, and determining timing and sequencing. Validate and improve the simulation model by comparing it to real-world data and making necessary modifications. Analyze the simulation findings to find areas for improvement and to enhance decision-making.

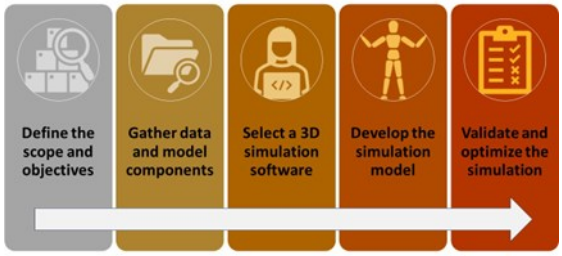


Figure 2: Building a 3D simulation model.

2.3. Visualization Techniques for Simulation

The use of virtual reality (VR), augmented reality (AR), and visual simulation technology allows to create a simulation model that can replicate real production scenarios in a safe environment (Xu et al., 2021). By inputting actual production data into the model, it is possible mimic on-site production in real-time. This technology helps us convert complex problems into more understandable ones (Xu et al., 2021). Furthermore, the simulation model provides a visual representation of production or manufacturing processes using terminal equipment (Abidi et al., 2020). They have further presented that by analysing and optimizing data, refining the mathematical model; and a closed-loop control and monitoring system can be created. This means that it enables to continuously monitor and adjust the production process based on the information provided by the simulation model.

A new strategy to generalizing the manipulation of industrial process simulation is explored (Abidi et al., 2020), where it is possible to visualize and engage with an industrial simulation model. It also enables industrial actors to visualize the operation of their factories which is based on virtual reality technology and behavioural programming of the production flow.

Xu et al., (2021) in their research describe the frame of visualization application, where they show the entire process starting from Cyber Physical System (CPS) to formulating the visualization platform. Within this VR and AR have been utilized for visualizing the created simulation model for the purpose of scheduling optimization and real-time monitoring. This results in a 3D recreation of industrial processes that depicts the production process that cannot be seen in depth on the 3D simulation of the production site (Xu et al., 2021).

2.4. Decision Making from simulation model.

Effective decision-making is crucial in industrial production processes as the decisions made during production can significantly impact efficiency, quality, and profitability (Yang et al., 2021). It is essential to have a clear understanding of the production processes and the consequences of different decisions on the outcomes. This is where the utilization of 3D simulation software and visualization techniques proves valuable (Xu et al., 2021). These tools enable manufacturers to create virtual representations of their production systems, visualizing the flow of materials, machines, and resources. Through simulation and visualization, decision-makers can assess different scenarios, identify potential issues, and make informed decisions before implementing changes in the physical world (Yang et al., 2021).

Figure 3 presents a conceptual model for decision making which has been designed to capture an expert's knowledge using simple and understandable elements, without relying on complex artificial intelligence or process representation (Garcia-Crespo et al., 2010). Instead, real-life situations have been focused, which occur during the execution of a manufacturing process, which consider facts (data), action (resources) and verification (checking the results from action are valid) through the process of making a decision.

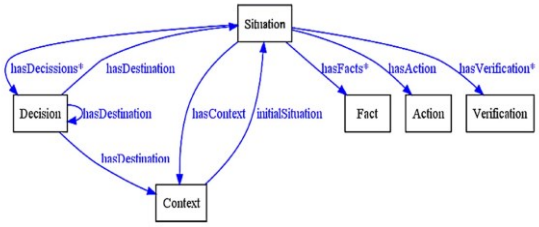


Figure 3: Decision Making Conceptual Model (Garcia-Crespo et al., 2010)

The aspect of decision making becomes important once there is an access to the real-time data through the created simulation model. Furthermore, Garcia-Crespo et al., (2010) in their research have presented this ontology for describing the automation of decision-making processes in the manufacturing process setting. In that, they have highlighted the importance of having a semantic representation of different manufacturing processes which further enable data representation strategy based on with the process of decision making is enabled.

3. Case Study

This section presents a 3D simulation model as a case study to showcase the application in real-world situation. This section further highlights the process of decision making and finally presents the results from the simulated model.

3.1. Case Description

Simulation modelling is essential for streamlining manufacturing procedures and increasing overall productivity in the area of industrial automation. Industrial systems can be accurately modelled and analysed to find opportunities for improvement using cutting-edge 3D simulation software. This study presents a simulation model of an industrial automation case in a sector of manufacturing with a focus on plastic injection-moulded basket manufacturing. The purpose of this study is to demonstrate how decision-making may be improved in the context of industrial production processes by using visualization and simulation approaches. The simulation model depicts a comprehensive representation of the production shopfloor, showcasing various interconnected components and processes.

Figure 4: 3D model of the production shopfloor

Figure 4 provides an overview of the shopfloor, offering a visual representation of the industrial automation case. The simulated model consists of two injection moulding machines, which are coupled by a conveyor system. This conveyor system facilitates the seamless transfer of the manufactured items (plastic baskets), between the machines. Subsequently, the baskets undergo a visual inspection process to identify any poor-quality elements. The conveyor system also enables the transportation of the approved baskets for further processing. To streamline the packing process, a robot is strategically placed between the two conveyor systems. This robot plays a vital role in efficiently packing the baskets into boxes. Once the baskets are securely placed in the boxes, they are seamlessly transferred to the palletization process. Here, another robot is responsible for picking up the boxes and organizing them on the pallets.

The palletization process incorporates its own conveyor system, ensuring a smooth flow of operations. Within the palletization process, the plastic wrapping and pallet stacking procedures take place, further enhancing the stability and protection of the palletized goods. Once the pallets undergo the stacking process, a stacker crane is employed to hoist them and place them in the designated warehouse for storage. This ensures efficient utilization of space and ease of access when retrieving the pallets for distribution or subsequent production stages. Through the integration of 3D simulation software, this case study showcases the visualization of an industrial automation scenario in the manufacturing industry. The subsequent sections of this paper will delve into the benefits, challenges, and decision-making aspects associated with the visualization of this simulation model.

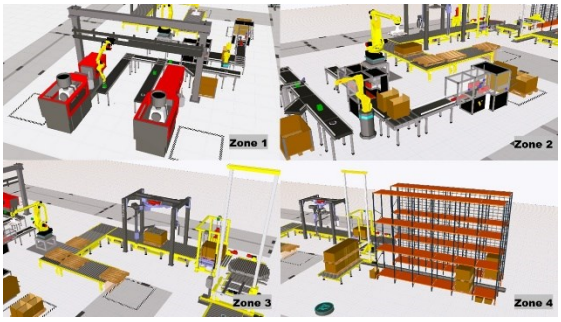


Figure 5: Different zones in the production shopfloor

The production shopfloor in the simulation model has been divided into four distinct zones: Zone 1, Zone 2, Zone 3, and Zone 4. Each zone encompasses specific functional procedures and components that contribute to the overall industrial production process. Figure 5 provides a visual representation of the layout and arrangement of these zones within the production shopfloor. The division of the production shopfloor into these zones and the incorporation of specific components and processes within each zone provide a comprehensive representation of the industrial production model.

Zone 1 serves as the initial stage of the production line, comprising two injection molding machines, a robot positioned overhead between the machines, and a conveyor transport system. The robot efficiently picks up the moulded baskets from the injection molding machines as the machine doors open. Subsequently, the robot places the baskets onto the conveyor belts, which then undergo a visual inspection process. During the visual inspection, any poor-quality baskets are identified and removed

from the conveyor belt line to ensure product quality. Moving on to Zone 2, this area is dedicated to the packaging system, which is constructed along the conveyor system. The packaging process begins with an automated machine that prepares folded boxes. These prepared empty boxes are then transported on the conveyor belt and stop near the robot. With the box in place, the robot organizes the baskets by picking and placing them in an organized sequential order within the box. Once the box is filled, it is transported to another automated machine, where it is sealed before being moved to the next conveyor belt.

Zone 3 represents the palletization process which involves a conveyor system. In this zone, the boxes filled with baskets are placed on pallets with the assistance of a robot. Once the pallets are filled with the boxes, the conveyor system transports them to an automated plastic wrapping station. At this station, the pallets undergo a plastic wrapping process, ensuring their stability and protection during transportation and storage. After the plastic wrapping is complete, the pallets are transferred through the conveyor line to an automated pallet stacking machine. This machine lifts the pallets and stacks them on top of one another, optimizing space utilization and facilitating efficient storage. Finally, Zone 4 is dedicated to the storage of the pallets in the warehouse. An automated stacker crane system is employed in this zone to hoist the pallets and arrange them in an organized manner. The design of the stacker crane system ensures that the warehouse storage racks are within the crane's reach, enabling seamless storage operations.

As a result, the simulation model described here provides a comprehensive picture of an industrial automation case in the manufacturing industry. The simulation model provides a full portrayal of the industrial production processes involved in making plastic injection-moulded baskets by dividing the production shopfloor into discrete zones and adding various components and processes. The 3D simulation software's representation of the production shopfloor allows for a clear knowledge of the interconnection and operation of each zone. This visualization helps in the identification of potential bottlenecks, areas for improvement, and the overall optimization of production efficiency, quality, and profitability. The simulation model demonstrates the importance of visualization and simulation approaches in improving decision-making in industrial manufacturing processes. Decision-makers can analyze the impact of many aspects on the production line, such as throughput, resource usage, and system performance, by modeling the different zones. This enables educated decision-making, which leads to better planning, optimization, and overall improvement of industrial manufacturing processes.

3.2. Decision Making

Creating a 3D simulation model for visualizing and optimizing industrial production processes necessitates a number of crucial decision-making elements. The creation of discrete zones on the production shopfloor gives an organized structure for analysing and improving each phase of the production line. The structure and arrangement of the production shopfloor is the first key decision-making component. The shopfloor is divided into particular zones, such as Zone 1, Zone 2, Zone 3, and Zone 4, allowing for a systematic approach to understanding and optimizing the many functional procedures involved. Choosing the best combination of resources, machinery, and equipment for each zone is a critical choice that affects the overall efficiency and productivity of the production line.

Decisions about the arrangement of injection moulding machines, the overhead robot, and the conveyor transport system were made within Zone 1. To ensure seamless operation and efficient material flow, factors such as machine capacity, cycle times, and coordination between machines and robots were taken into consideration. Moving on to Zone 2, the decision-making process constituted of creating an efficient packaging system along the conveyor system. Choosing the right automated machine for folding boxes, deciding the positioning and sequencing of baskets by the robot, and ensuring effective box sealing are all important factors that affect the overall packing process. Decisions in Zone 3 revolve around the palletization process. The selection and location of the robot for palletizing the boxes, optimizing the conveyor system for efficient box transfer, and constructing the plastic wrapping station to ensure secure and stable pallets for transit and storage are all important considerations. Finally, in Zone 4, the decision-making process revolves around pallet storage in the warehouse. Choosing the right automated stacker crane system, customizing the crane's reach and movement, and correctly organizing the storage racks are all critical decisions in order to maximize space utilization and streamline storage operations.

It is essential to examine elements such as system throughput, resource usage, material flow optimization, and overall production efficiency throughout the decision-making process. Simulation modelling provides for the testing of numerous scenarios, the evaluation of the impact of various choice alternatives, and the identification of potential bottlenecks or areas for improvement.

3.3. Simulation Results

During the simulation, a total run time of 8 hours was chosen to align with the duration of a typical daily shift in the industrial production environment. This time frame allows for a comprehensive evaluation of the production processes and the performance of the simulation model. In the simulation model, three robots were strategically positioned at the junctions between the zones to facilitate the smooth transfer of baskets and optimize the overall material flow.

The first robot, Robot 1, was placed overhead, directly above the two injection-moulding machines in Zone 1. This robot plays a crucial role in picking up the moulded baskets from the machines and transferring them to the conveyor belts. Robot 2 was positioned between Zone 1 and Zone 2, acting as a key component in the transition from the injection moulding process to the packaging system. It assists in organizing the baskets and placing them in the boxes in a sequential order. Similarly, Robot 3 was placed between Zone 2 and Zone 3, facilitating the movement of the filled boxes from the packaging system to the palletization process. This robot plays a vital role in placing the boxes on pallets, ensuring efficient palletization. Figure 6 provides valuable insights into the utilization of these three robots throughout the 8-hour simulation run. The utilization is represented as a percentage, indicating the proportion of time each robot was actively engaged in performing its designated tasks.

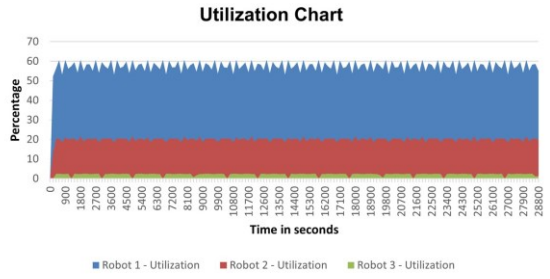


Figure 6: Robot percentage utilization

These presented results allow for a deeper understanding of the robots' workload distribution and their efficiency in supporting the production processes. By evaluating the utilization percentages, it becomes possible to identify potential bottlenecks or areas of improvement within the production line. For instance, if a robot's utilization is consistently high, it may indicate a need for additional resources or process optimization to alleviate the workload. On the other hand, low utilization may suggest under-utilized capacity that can be optimized for increased productivity. The utilization analysis of the robots provides valuable insights into the operational efficiency and resource allocation within the simulated industrial production system. These findings can guide decision-making processes to enhance the overall performance and productivity of the production processes.

The simulation model incorporates various process nodes that are programmed to operate within specific flow groups. In this particular simulation, there is a single flow group representing the production of a specific product type, which in this case is baskets. Each process node within the model has the capability to assume different statistical process states, including idle, busy, and blocked.

To establish the interconnection between different zones, four process nodes were selected in this simulation model. These process nodes serve as the critical links between the zones and play a crucial role in the overall flow of the production process. Figure 7 visually represents the state results obtained from these zonal process nodes. The results are presented as percentages, reflecting the distribution of different process states throughout the entire production or simulation run time, which was set to 8 hours in this case.

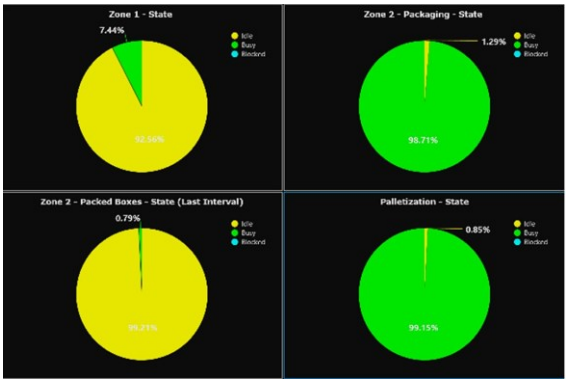


Figure 7: Idle, busy, and blocked States

Starting with the Zone 1 state process node, it represents the initial stage where the baskets arriving on the conveyor belt are picked up by Robot 2 and subsequently placed in the Zone 2 (Packaging) process node according to the programmed assembly sequence. The Zone 2 (Packed Boxes) process node captures the state of the final packed boxes. Here, Robot 3 is responsible for lifting the packed boxes and placing them onto the pallet, which is represented by the Zone 3 (Palletization) process node. Robot 3 follows a specific assembly sequence while arranging the boxes on the pallet. It is noteworthy that each pallet accommodates a total of 8 boxes, organized in a 2x2x2 stack configuration.

By analysing the state results from these zonal process nodes, it becomes possible to gain insights into the efficiency and performance of the production process. The percentages offer a quantitative assessment of the process states, indicating the proportion of time spent in each state throughout the 8-hour simulation run. These results

enable the identification of potential bottlenecks, areas of improvement, or process optimization opportunities within the production flow.

4. Analysis and Discussion

The simulation model focuses on improving decision-making in industrial production processes, in the context of plastic injection-moulded basket manufacturing. The simulation model provides a visual representation of the production shopfloor, encompassing various related components and processes, using 3D simulation software Visual Components®. This was further tested in Virtual Reality (VR) environment (Figure 8) to analyse the simulation model and to initiate the decision-making process within the different zones which are representing different stages of the production line. The simulation model's decision-making process included identifying the structure and layout of the production shopfloor, selecting the necessary resources and equipment for each zone, and optimising aspects such as material flow and system throughput. Additional VR analysis enabled the process of deciding how much space to be available so that operators can use it for maintenance.

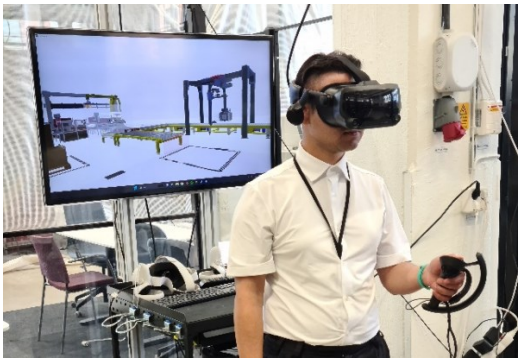


Figure 8: Simulation model testing in VR

Simulation results show the utilization of robots positioned in different zones, allowing for the identification of potential areas for improvement. It was observed from the robot utilization chart (Figure 6), that the utilization percentage for Robot 2 was less and Robot 3 was much lesser. This tells us that there is a possibility for further expanding Zone 1 with more injection-moulding machines. The analysis of process states within the zones provided insights into the efficiency and performance of the production process. Similarly, it was also observed in Zones 1 and 2, the idle time amounted to more than 90% of the total simulation run time. Because of this, it is possible to advise the business management that the current process will be able to support any future investments in additional injection moulding machines in the case concerning increasing the production capacity.

Therefore, to answer the research question, the use of 3D simulation software and virtual reality (VR) in visualizing industrial production processes provides useful insights to decision-makers for optimizing production efficiency. It enables informed decision-making by providing a comprehensive picture of the production flow, resource utilization, and system performance.

Decision-makers can increase the overall efficiency and profitability of manufacturing processes by identifying areas for improvement and resolving potential bottlenecks. The validation of the simulation model was done during the development phase through the immediate testing of simulation animation results in the 3D world environment within the software. This validation was in the form of correlating and checking if the flow components were in accordance to the set requirements.

5. Conclusion

In conclusion, this study highlighted the significance of 3D simulation modelling and visualization techniques in improving decision-making and optimizing production efficiency in industrial manufacturing processes, with a specific focus on plastic injection-moulded basket manufacturing. The utilization of 3D simulation software, coupled with VR testing, allowed for a comprehensive and immersive representation of the production shopfloor, showcasing interconnected components and processes.

Decision-makers can examine and comprehend the production flow, resource allocation, and system performance with the help of the simulation model. Additionally, the developed model is flexible in order to accommodate future modifications. By examining the utilization of robots in different zones, potential areas for improvement can be identified, leading to enhanced resource utilization and increased production efficiency. The analysis of process states within the zones further contributes to identifying bottlenecks and optimization opportunities within the production process.

By offering visual context, dynamic interaction, and the capacity to test possibilities, 3D simulations and VR technologies transform shopfloor decision-making. They uncover complicated relationships, possible bottlenecks, and ergonomic difficulties that statistics overlook. These systems detect problems early, improve communication, and provide a visual knowledge of shopfloor processes. They alter structure, throughput, and efficiency decisions by displaying the shopfloor layout, equipment, and processes. Unlike traditional numerical analysis, 3D

simulations and VR provide decision-makers with immersive insights, allowing them to make educated decisions that lead to optimised industrial production processes. The findings from the simulation model and visualization techniques demonstrate the importance of informed decision-making in industrial manufacturing. Decision-makers can leverage the insights gained from the simulation model to make educated decisions regarding the structure and arrangement of the production shopfloor, selection of appropriate resources and equipment, and optimization of material flow and system throughput.

Acknowledgment

I would like to thank MITC AB, Eskilstuna for providing me with the resources which were valuable for conducting this research.

References

Abidi, M.-A., Chevaillier, P., Lyonnet, B., Kechiche, M., Baert, P., & Toscano, R. (2020). *How to Create a New Generation of Industrial Processes Simulation by Coupling Simulation Tools with VR Platforms*.

Atmakuri, S., Upadhya, G., Wang, C. M., Chandra, U., & Paul, A. J. (1993). The role of visualization in process simulation. *JOM*, *45*(10), 21–24. https://doi.org/10.1007/BF03222455

Chawla, R., & Banerjee, A. (2001). A virtual environment for simulating manufacturing operations in 3D. *Proceeding of the 2001 Winter Simulation Conference (Cat. No.01CH37304)*, *2*, 991–997. https://doi.org/10.1109/WSC.2001.977404

Garcia-Crespo, A., Ruiz-Mezcua, B., Lopez-Cuadrado, J. L., & Gomez-Berbis, J. M. (2010). Conceptual model for semantic representation of industrial manufacturing processes. *Computers in Industry*, *61*(7), 595–612. https://doi.org/10.1016/j.compind.2010.01.004

Molenda, P., Jugenheimer, A., Haefner, C., Oechsle, O., & Karat, R. (2019). Methodology for the visualization, analysis and assessment of information processes in manufacturing companies. *Procedia CIRP*, *84*, 5–10. https://doi.org/10.1016/j.procir.2019.04.291

Wang, J., Phillips, L., Wu, B., Moreland, J., & Zhou, C. (2015). *Simulation and Visualization of Industrial Processes in Unity*.

Xu, T., Song, G., Yang, Y., Ge, P., & Tang, L. (2021). Visualization and simulation of steel metallurgy processes. *International Journal of Minerals, Metallurgy and Materials*, *28*(8), 1387–1396. https://doi.org/10.1007/s12613-021-2283-5

Yang, T., Yi, X., Lu, S., Johansson, K. H., & Chai, T. (2021). Intelligent Manufacturing for the Process Industry Driven by Industrial Artificial Intelligence. *Engineering*, *7*(9), 1224–1230. https://doi.org/10.1016/j.eng.2021.04.023

Optimal indoor temperature flexibility for thermal peak shaving in buildings connected to the district heating network

Mathilda Cederbladh[a,*], August Dahlberg[a], Stavros Vouros[a], Konstantinos Kyprianidis[a], Costanza Saletti[b], Mirko Morini[b]

[a] *Mälardalen University,* [b] *University of Parma*
mathilda.cederbladh@gmail.com

Abstract

Buildings are currently non optimally controlled, using a weather compensation controller that depends only on external temperature. A rich amount of real-time data is available and can be used for better control. This work is focused on developing a general and dynamic model for utilizing the building as an energy storage for a peak-shaving control strategy. A dynamic grey-box model is developed using industry standard operators data from a multifamily building, Building A, located in Västerås, Sweden. The training period is set to 408 hours, and the prediction horizon to 48 hours. The model is verified in 4 steps: prediction ability on the historic data, parametric verification on the time constant, simulation of heat supply separated from the historic data and model generality by implementing the model on a second multifamily building, Building B. The modelling errors over a two-month simulated period are 8 % for Building A and 9 % for Building B. To demonstrate the utilization possibilities, an optimizer is constructed to evaluate a peak shaving control strategy. Different flexibilities for the indoor temperature have been examined with a range yielding heat load peak shaving between 30 to 45%. Flexibility paves the way for improvement in pricing models for the heating sector. This work demonstrates the potential of utilizing building heat storage capacity to reduce peak consumption and costs.

1. Introduction

Currently, District Heating (DH) substations operate in sub-optimal conditions due to a lack of information about the supplied buildings, their future demand, and the operating parameters. The rich amount of real-time data available from new sensors implies saving potential if made available to the energy providers and buildings managers. Utilizing building thermal inertia as a short-term storage is a cheap and viable technology (Kensby et al., 2015), the concept consists of overheating or underheating the building. When overheating or underheating the building a change between the set indoor temperature and the actual indoor temperature occurs. This results in a divergence from the set temperature, and it is that temperature difference that functions as the energy storage in the building (Ståhl, 2009). By utilizing the internal heat transfer in buildings as heat storage the supply need can be reduced, assuming that the producers have knowledge of the relevant storage data. One of the main constraints in utilization lies in the comfort requirements of the occupancies (Renström et al., 2021). This work is focused on developing a general and dynamic model for utilizing buildings as energy storage for a peak-shaving control strategy. The work aims to determine how stored heat in buildings can be modelled using industry standard data streams. Furthermore, the work investigates the potential in controlling a building's heating system with consideration to stored heat and how a flexible indoor temperature affects different aspects of building thermal control.

2. Methodology

2.1. Problem setup

The building used for model development is a 9-storey multifamily building, called Building A, located in Västerås, built in 2017.

The thermal dynamics of the building consist of multiple different heat sources and heat losses from building components. The heat sources consist of two sources; heat supplied from the DH system, and heat delivered from unmeasured sources (passive heating). The main sources of heat loss are through the building envelope and the ventilation. The input data was originally supplied by the local DH company Mälarenergi (primary side) and a local landlord Mimer (secondary side). The data has a time step of one hour and consists of 1501 data points between 2019-12-01 and 2020-02-02. The temperature has been taken as an average over all the individual apartments to give an average temperature for the building itself, therefore the standard deviation on the indoor temperature is also given. The outdoor temperature was measured with a sensor installed on the building. Supply temperature, water mass flow and return

temperature were all measured at the building's main heat exchanger.

The heat supply (calculated based on the supply-, return temperature and the mass flow from the main heat exchanger), average indoor temperature, and external temperature are used in this work. The data is presented in Fig. 1.

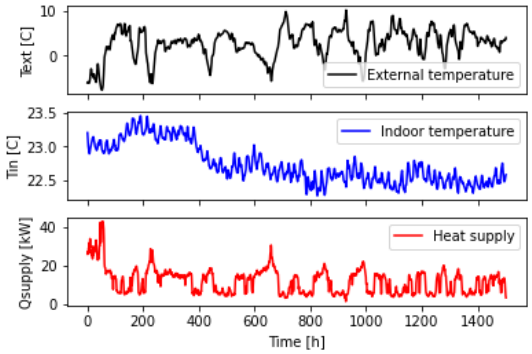


Figure 1: Historical data for Building A

2.2. Development of Building Model

The finalized model consists of three models that combine the strength of them all. The first is called First Order Thermal Model (FOTM) and is based on simple 1R1C models as described by Harb et al. (2016) or Monghasemi et. al (2022). The second is called Degree Day Model (DDM) and is a further development of the 1R1C model using the degree day method by Tabatabaei et al. (2017). The third is called Time Constant Model (TCM) and is based on Antonopoulos & Koronaki (2000). The cooperation between the different models is illustrated in Fig.2.

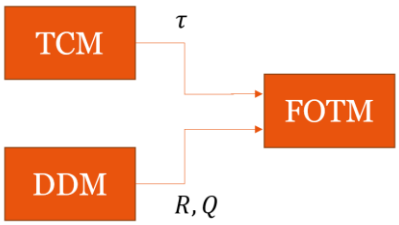


Figure 2: Model flow chart

FOTM is developed as an 1R1C model to ensure robustness due to the lower complexity. The typical 1R1C model is described with Eq. 1.

$$C\frac{dT}{dt} = R(T_{ext} - T_{in}) + Q_{supply} \quad (1)$$

FOTM includes an additional parameter to act as the heat from other sources then the heat delivered from the building's substation, $Q_{passive}$. This simulates heat from occupancies, electric appliances, solar radiation etc. The FOTM model is described with Eq. 2.

$$C\frac{dT}{dt} = R(T_{ext} - T_{in}) + Q_{supply} + Q_{passive} \quad (2)$$

The DDM is set up as a function called from the FOTM, it utilizes Eq. 2 with $\frac{dT}{dt} = 0$, resulting in Eq. 3**Error! Reference source not found.**.

$$R(T_{ext} - T_{in}) - Q_{passive} = Q_{supply} \quad (3)$$

Curve fitting (based on least squares) is then used to estimate $R_{passive}$ and $Q_{passive}$. The steady state heat loss of the building is represented by R. There are always some daily variations in the indoor temperature, but the general trend of the indoor temperature must be steady for R determination. $Q_{passive}$ is determined as an average of the passive heating during the training period and will therefore give an imperfect estimation on the hourly passive heating in the building.

TCM is also set up as a function called by FOTM. The equation used is Eq. 4 where the parameters C_{eff} & U are determined by optimizing for them using least squares from the SciPy optimize library with Eq. 5 as the cost function.

$$T_{in}(t) = T_{ext} - \left[T_{ext} - T_{in}(0) + \frac{Q_{supply}}{U}\right] e^{\left(-\frac{tU}{C_{eff}}\right)} + \frac{Q_{supply}}{U} \quad (4)$$

$$F(X) = T_{in,pred} - T_{in,real} \quad (5)$$

Given the resulting C_{eff} and U for the training period the mean of the values is taken to calculate the time constant according to Eq. 6.

$$\tau = C_{eff}/U \quad (6)$$

The time constant is then used to calculate a C based on the R value from DDM. In contrast to DDM, TCM requires variance in the indoor temperature to determine the effective heat capacity, C_{eff}.

The model detects the trend of the indoor temperature, thereby not capturing daily variations. The model is a dynamic model with accuracy dependent on the accuracy of the temperature data, as the temperature sensor most likely has the highest uncertainty. The temperature sensor is unknown; however a typical range is ± 1 K according to manufacturer specification data.

2.3. Verification of building model

The first verification step is to determine the prediction ability by implementing the model on the entire available data set and observing the model's

ability to predict the indoor temperature for each time step over different prediction horizons. The model must replicate the trend rather than the actual values to avoid overfitting to variance caused by measuring the input data in the air. The next step is to verify the parameters. This is done by observing the resulting values of the time constant and comparing it to expected values from the literature, as presented in Tab. 1.

Table 1: Time constants presented by Johra et al. (2019)

Time constant τ [h]	Light	Medium	Heavy
1980's house	9	49	181
Passive house	135	169	626

The heat is then simulated using the historical outdoor temperature without access to the historical heat data. The generality of the model is also verified by implementing the model on a different building, Building B. The data available for Building B is of a similar character to that of Building A.

2.4. Optimizer

To highlight the utilization possibilities of using the building as a heat storage, an optimizer is developed to evaluate the future heat supply. It is developed in Gekko, which is an optimization library in python (Beal et al., 2018). The optimizer is based on a strategy presented by Saletti et al. (2021) that focuses on minimizing the variations in heat supply by controlling for the derivative of the heat supply.

To minimize the total variation in the heat supplied (Q_{supply}) the derivative squared is minimized to ensure that only positive values occur in the objective, this will assist the optimizer in flattening out the heat demand. The control objective (Q_{supply}) is controlled by manipulating the indoor temperature (T_{in}) within a set interval. By setting a fixed interval for the indoor temperature within the comfort interval, the comfort is still maintained. The optimization problem is stated below.

$$Objective: \quad \min\left(\sum_{i=1}^{i=forecast} dQ_{supply_i}{}^2\right)$$

$$Constraints: \quad T_{in_{min}} \leq T_{in_i} \leq T_{in_{max}}$$

$$Q_{supply_{min}} \leq Q_{supply_i} \leq Q_{supply_{max}}$$

$$Q_{supply_i} = \left(T_{in_{i+1}} - T_{in_i}\right)C - R\left(T_{ext_i} - T_{in_i}\right) - Q_{other}$$

Before the model can be utilized for peak shaving, the training period and the prediction horizon need to be determined by parametric analysis to yield the most accurate results. The training period is set to 408 hours/data points, and the prediction horizon is set to 48 hours. Different flexibilities in the indoor temperature are examined from Tin=22°C ± 0.25 to Tin=22°C ± 2.00. The baseline temperature is set to 22 °C to ensure that the indoor temperature is always within the Swedish health agencies recommended comfort interval (Folkhälsomyndigheten, 2014).

2.5. Normalized Economics

Currently, economically motivated peak shaving for customers is not the norm at the local DH company. The customers have the option to select a "baseload consumption" with a fixed price in SEK⁄kW,year and a "peak consumption" with a different fixed price in SEK⁄kW,year. But they can also allow the company to choose, and then no fixed cost is added if the consumption increases above the baseload level (Landelius & Åström, 2019). To highlight the economic benefits in peak shaving from a customer perspective a normalised economic analysis is developed. An initial baseload is calculated based on the average outdoor temperature for December 2019 using Eq. 3. The different parameters are estimated on an average of the first 408 data points.

The economic savings are then calculated by integrating the curves that exceed the baseload and comparing the integrals for the historic case and the new peak shaved case, as shown in Fig. 3.

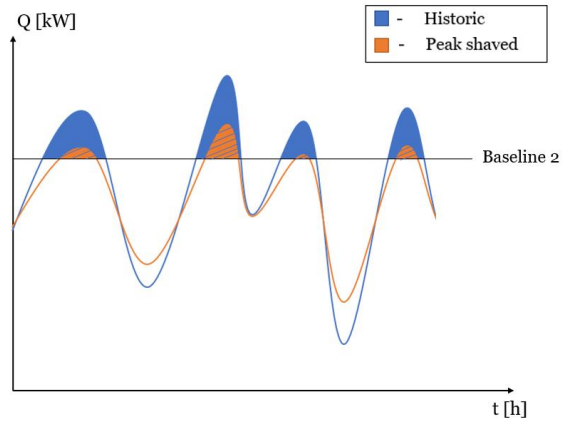


Figure 3: Normalized Economic Analysis

The savings are presented as a percentage and are calculated according to Eq. 7.

$$Savings = 100 - \frac{\int Peak_{new}}{\int Peak_{historic}} 100[\%] \qquad (7)$$

The baseload is then varied to illustrate the different savings achieved depending on the baseload level.

3. Results

3.1. Building model verification

The mean and max error over the entire prediction horizon for each time step when choosing the optimal prediction horizon and training set length is plotted in Fig. 4.

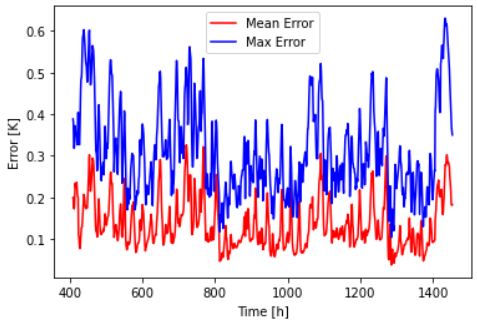


Figure 4: Mean and maximum error for Building A

The maximum error is 0.63K and the mean error of about 0.14K for Building A.

The heat capacity of Building A is 108 $[\frac{Wh}{Km^2}]$ respectively 124 $[\frac{Wh}{Km^2}]$ for Building B. The average time constant over the entire data set for Building A is 209 hours, and for Building B 120 hours. Compared to the values given in Tab. 1., these are reasonable.

As illustrated in Fig. 5. & Fig. 6., the simulated heat supply and indoor temperature generated by the model are reasonably well correlated to the historical data, as shown in Tab. 2.

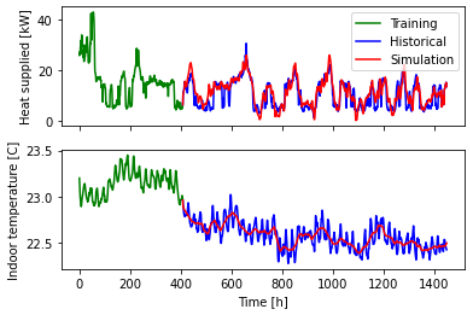


Figure 5: Verification of model for Building A

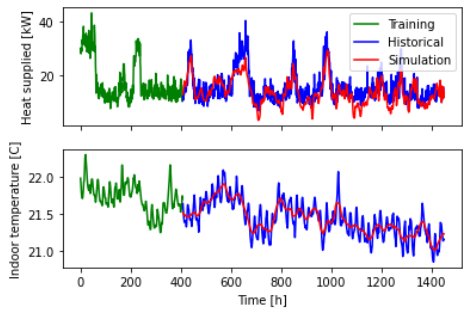


Figure 6: Verification of model for Building B

Table 2: Verification results: comparison between simulated and historical data for heat supply

	Building A	Building B
Correlation coefficient	85%	77%
RRMSE	8%	9%

The fact that the simulated heat supply and the historical heat supply do not match each other perfectly is not of concern, since it is the trend that is of interest.

3.2. Utilization potential

Fig. 7. shows the cumulative distribution of the heat supplied during the simulation period.

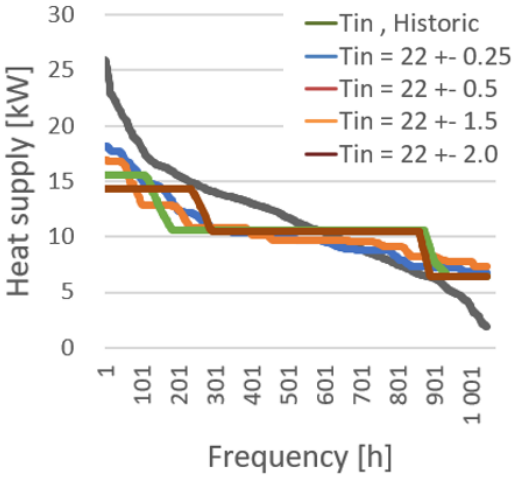


Figure 7: Load duration curve for peak shaving in Building A

The frequency describes how often (how many hours) the corresponding heat supply is reached for each investigated frequency. In general terms, by increasing the flexibility of the indoor temperature, less variation in the supplied heat and a lower peak power is achieved.

In Tab. 3. the results for different temperature flexibilities are presented. It shows that in general when implementing the peak shaving control strategy there is also an overall reduction in heat supply. However, this is mainly due to a reduction in the set indoor temperature as the simulated historical data has an average indoor temperature of 22.6 °C. Therefor it can't be concluded that utilizing peak shaving results in energy savings, it can however reduce emissions. If utilized by DH companies, peak shaving could reduce the need for a fossil fuel boiler during peak hours and thereby reduce overall emissions and the dependency on expensive and harmful fossil fuels.

Table 3: Results from peak shaving compared to the simulated historical data

Temperature flexibility [°C]	Peak power decreased	Energy consumption decreased	Average indoor temperature [°C]
±0.25	30	10	22
±0.50	35	10	22
±0.75	35	12	21.9
±1.00	36	9	22
±1.25	38	14	21.9
±1.50	40	7	22.2
±1.75	42	7	22.3
±2.00	45	6	22.3

As seen in Tab. 3. a temperature flexibility of ±1.25 $°C$ generates the most energy savings, however in a survey by Renström et al. (2021) including 88 respondents it was found that approximately 40 % of respondents thought they would not be affected by an variable indoor temperature of ±1 $°C$ whilst approximately 60 % thought they would be negatively affected by a variable temperature of ±1.5 $°C$. This might intel that even though the comfort impact will be minimal, it will be difficult to convince consumers to use the developed approach with higher variability. Renström et al. (2021) even found that 20 % of respondents believed they would be negatively affected already at ±0.5 $°C$.
To determine the appropriate flexibility a normalized economics analysis is done, the results are presented in Fig. 8.

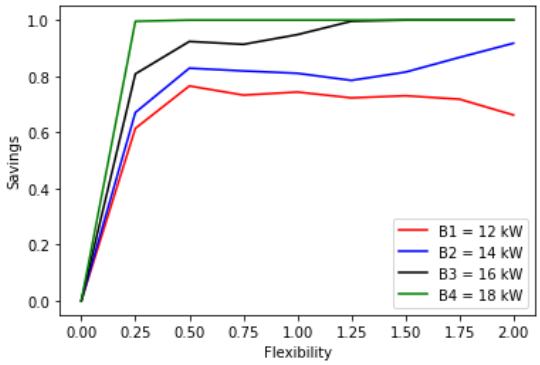


Figure 8: Normalized economics analysis results

As the percentage of the available area is made up of the difference between the historical peak and the shaved peak, the magnitude of the baseload is directly correlated to the possible savings. From the combined knowledge presented in Fig. 8.**Error! Reference source not found.** A higher baseload results in larger savings, this is demonstrated clearly in Fig. 9.

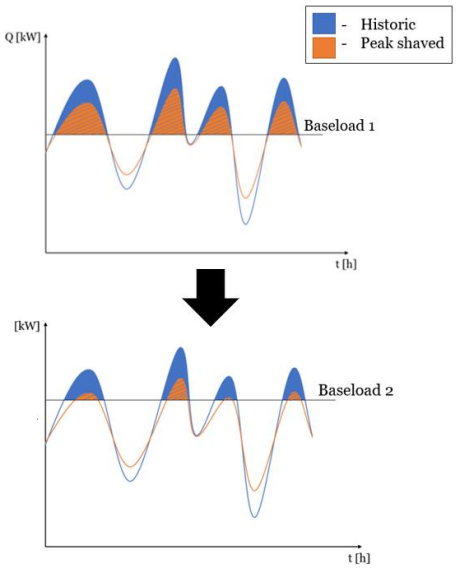


Figure 9: A moving baseload's impact on savings

In Fig. 9., one can determine that choosing an appropriate baseload is crucial when it comes to the economic benefits of peak shaving. The savings in the fixed cost of the peak consumption depends on how much of the peaks are above the baseload. Therefore, the higher baseload reaches savings of 100% since the new peak is below the baseload. In the lower baseloads, where the new peak never is below the baseload, the preferred choice of temperature flexibility is $\pm 0.5°C$. Due to this a temperature flexibility of $\pm 0.5°C$ is chosen as the preferred interval. Fig. 10. and Fig. 11. shows the potential of peak shaving by allowing a flexibility in the indoor temperature with ±0.5 $°C$ for Building A and B.

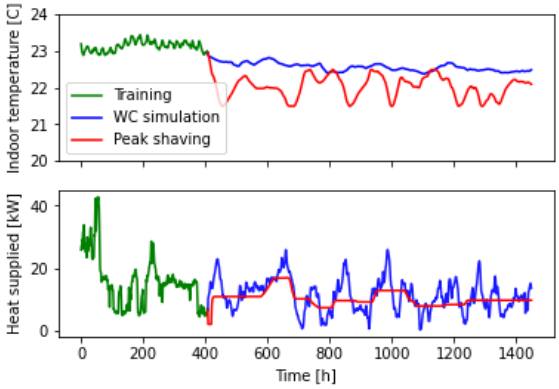


Figure 10: Peak shaving, Tin = 22°C ±0.5, prediction horizon = 48, Building A

Fig. 10. shows that the highest peak in the heat supply during the simulation decreased from 26 kW to 16 kW for Building A, and the total energy consumed during the simulation decreased from 12 100 kWh to 11 000 kWh.

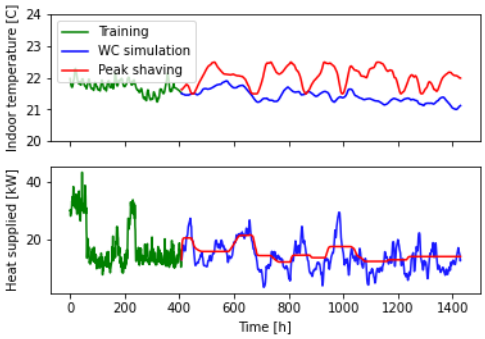


Figure 11: Peak shaving, Tin = 22°C ±0.5, prediction horizon = 48, Building B

For Building B as shown in Fig. 11. the highest peak was decreased from 29 kW to 22 kW, and the total energy consumed was increased from 14 700 kWh to 16 000 kWh.

4. Discussion and conclusions

4.1. Discussion

The FOTM and the constructed optimizer is a general model with a low execution time. Since the model was developed using the data from Building A and then also tested on data from Building B, the generality requirement was met. Execution time has not been an issue in the tests (448 seconds for a simulation period of 1085 hours). However, there are some concerns regarding the reliability when using online adaptation, since there are certain rules that must be considered for parameter estimation, which have not been developed. This have mainly been the parameter estimation of R and $Qpassive$. When estimating these parameters, the most important part is that the indoor temperature is stable and does not vary over time. When adapting peak shaving it is never stable. Vice versa if the indoor temperature is kept steady $Ceff$ is hard to estimate since it requires change within the indoor temperature to be determined. Examining how these issues might be solved belongs to the future works. For the results presented the parameters $R_{passive}$ and $Ceff$ has been determined on the training set and the assumed constant for the entire simulation period. This gives reliable results for the available data as shown.

By utilizing a control strategy similar to the one suggested in this work there is a possibility of increasing the amount of electricity produced in a CHP plant. A reduction in heat demand from the consumers yields a larger portion of the produced heat at the plant available for electricity production. Currently, most renewable sources cannot keep up with the higher electricity demand during the winter where DH companies normally can't produce any electricity due to the high heat demand. By decreasing the heat demand and increasing the electricity production the revenue can be greatly increased.

The authors suggest that a new subscription format is produced in which the customers subscribe to a certain comfort interval rather than a certain heat flux. The authors also believe that customer participation and engagement should be an integral part of future business models to ensure customer satisfaction. The authors also suggest adding a safety margin in the subscription range.

4.2. Conclusions

In this work a data driven physics-based model has been produced which directly quantifies the steady state heat loss, the heat capacity, the time constant and the passive heating. These parameters are essential for determining a building's thermal storage ability. The model has a RRMSE of 8% for Building A and 9% for Building B. By having access to the thermal dynamics of a buildings storage potential the buildings heating system can be controlled from a peak shaving perspective which utilizes over- and underheating to charge and discharge the building around peaks in heat demand generated by changes in the external temperature. Utilizing a peak shaving control strategy has been shown to generate savings in energy consumption of up to 14 % and 45 % in peak consumption (depending on the set indoor temperature and allowed flexibility). The authors suggest allowing the customers to choose their preferred flexibility to ensure their comfort, but also point out that a beneficial control can be found at an indoor temperature of 22 ±0.5 °C.

There are four aspects to be considered when allowing for a flexible indoor temperature, comfort, energy consumption, peak shaving, and economics. All aspects are correlated and to find an optimal strategy, compromises must be made. The authors have shown that economic savings and peak shaving can be achieved by allowing for small variations in the indoor temperature to the detriment of comfort and in some cases energy consumption.

Acknowledgment

The authors would like to acknowledge Mälarenergi AB for partially financing this work and providing operational data for the substation. Mimer AB is acknowledged for kindly providing the building data. Particular acknowledgements go to Joacim Sundqvist for his continuous support. The work has been also partially financed by the project DISTRHEAT (Digital Intelligent and Scalable Control for Renewables in District Heating Networks) under the umbrella of ERA-Net Smart Energy Systems.

References

Antonopoulos, K. A., & Koronaki, E. P. (2000). Thermal parameter components of building envelope. Applied Thermal Engineering, 20(13), 1193–1211. https://doi.org/10.1016/S1359-4311(99)00090-3

Beal, L., Hill, D., Martin, R., & Hedengren, J. (2018). GEKKO Optimization Suite. Processes, 6(8), 106. https://doi.org/10.3390/pr6080106

Folkhälsomyndigheten. (2014). Folkhälsomyndighetens allmänna råd om temperatur inomhus. 8.

Harb, H., Boyanov, N., Hernandez, L., Streblow, R., & Müller, D. (2016). Development and validation of grey-box models for forecasting the thermal response of occupied buildings. Energy and Buildings, 117, 199–207. https://doi.org/10.1016/j.enbuild.2016.02.021

Johra, H., Heiselberg, P., & Dréau, J. L. (2019). Influence of envelope, structural thermal mass and indoor content on the building heating energy flexibility. Energy and Buildings, 183, 325–339. https://doi.org/10.1016/j.enbuild.2018.11.012

Kensby, J., Trüschel, A., & Dalenbäck, J.-O. (2015). Potential of residential buildings as thermal energy storage in district heating systems – Results from a pilot test. Applied Energy, 137, 773–781. https://doi.org/10.1016/j.apenergy.2014.07.026

Landelius, E. and Åström, M., (2019). DISTRICT HEAT PRICE MODEL ANALYSIS: A risk assesment of Mälarenergi's new district heat price model. Mälardalen University.

Monghasemi, N., Vouros, S., Kyprianidis, K. and Vadiee, A., (2022) A non-linear gray-box model of buildings connected to district heating systems. Energy Proceedings Vol 29.

Renström, S., Ünlütürk, B., Nyström, S., & Monsberger, C. (2021). END-USER FLEXIBILITY POTENTIAL. Report V.1.0, Flexi-Sync.

Saletti, C., Zimmerman, N., Morini, M., Kyprianidis, K., & Gambarotta, A. (2021). Enabling smart control by optimally managing the State of Charge of district heating networks. Applied Energy, 283, 116286. https://doi.org/10.1016/j.apenergy.2020.116286

Ståhl, F. (2009). Influence of thermal mass on the heating and cooling demands of a building unit. 172.

Tabatabaei, S., van der Ham, W., C. A. Klein, M., & Treur, J. (2017). A Data Analysis Technique to Estimate the Thermal Characteristics of a House. Energies, 10(9), 1358. https://doi.org/10.3390/en10091358

Sustainability analysis and simulation of a Polymer Electrolyte Membrane (PEM) electrolyser for green hydrogen production

Jordi Béjar-Rabascall[a] Gaurav Mirlekar[b,*]

[a]*Escola Tècnica Superior d'Enginyeria Industrial de Barcelona (ETSEIB);Universitat Politecnica de Catalunya (UPC); Barcelona 08028, Spain*
[b]*Department of Electrical Engineering, Information Technology and Cybernetics; University of South-Eastern Norway (USN); Porsgrunn 3918, Norway*
[*]gaurav.mirlekar@usn.no

Abstract

In recent years, green hydrogen has emerged as an important energy carrier for future sustainable development. Due to the possibility of not emitting CO_2 during its generation and use, hydrogen is considered a perfect substitute for current fossil fuels. However, a major drawback of hydrogen production by water electrolysis, supplied by renewable electricity, is its limited economic competitiveness compared to conventional energy sources. Therefore, this work focuses on analyzing the sustainability of a green hydrogen production plant, not only considering its environmental parameters, as well as its economic, energy and efficiency parameters. The polymer electrolyte membrane (PEM) is selected as the most promising method of green hydrogen production in the medium and long term. Subsequently, a small-scale production plant is simulated using chemical process simulation software to obtain key data for computing a set of sustainability indicators. The selected indicators are based on the Gauging Reaction Effectiveness for the Environmental Sustainability of Chemistries with a Multi-Objective Process Evaluator (GREENSCOPE) methodology and are used to compare the sustainability of the simulated PEM plant with alkaline water electrolysis (AWE) plant. Finally, the process is scaled-up to analyze the feasibility of the simulated PEM system and validated against data to determine the operation of the electrolyser at a large production scale.

1 Introduction

Green hydrogen is typically obtained via water electrolysis that uses renewable power source to generate hydrogen from water. This method allows for the production of hydrogen without any CO_2 emissions (Younas et al., 2022). The key element of water electrolysis is electrolyser, a device where direct electric current is applied resulting in two chemical reactions: one at the cathode that produces hydrogen, Equation (1), and the other at the anode, Equation (2), that generates oxygen (Noussan et al., 2021). The polymer membrane-based electrolyser, also known as PEM, is the most suitable choice for handling power variations due to its ability to quickly adapt to changes within seconds, unlike other types of electrolysers based on alkaline water electrolytes (AWE) that require minutes. In addition, PEM can operate at higher pressures than other electrolysers, reducing the sub-sequence compression stages (Wang et al., 2022).

Cathode reaction:

$$2H^+ + 2e^- \rightarrow H_2(g) \tag{1}$$

Anode reaction:

$$H_2O(l) \rightarrow \frac{1}{2}O_2(g) + 2H^+ + 2e^- \tag{2}$$

Despite the fact that green hydrogen offers an environmentally cleaner solution to reduce society's dependence on fossil fuels, the reality remains that most of the world's hydrogen production, approximately 96% in 2021, still comes from methods that emit greenhouse gases, commonly known as grey and brown hydrogen (IRENA, 2021). The reason for the present scenario is that current green methods, such as PEM or AWE water electrolysis supplied by renewable power, are not economical, which makes them less competitive than those based on fossil fuels (Younas et al., 2022). This article therefore proposes a method for quantifying the sustainability assessment of the hydrogen production process, creating tools to evaluate of green hydrogen production methods. In addition, it is considered important to explore cost reduction measures during sustainability analysis of the process. Therefore, the advancement of process simulation is seen as a viable approach to minimise design costs. However, a challenge arises due to the limited availability of larger scale simulation models. To address this limitation, the simulated process is scaled-up to

evaluate the performance of the electrolyser in large-scale production scenarios.

2 Sustainability

By examining existing literature, this analysis assesses the present state of green hydrogen in terms of its environmental impact, efficiency, economic viability, and energy performance.

Environmental

Green hydrogen is considered as an environmentally friendly, carbon-neutral energy carrier. The contribution to climate change is minimal, as only oxygen is emitted during the production process. However, the environmental impact of hydrogen is not zero, the type of renewable energy source used, the origin of the water for the electrolysis process and the residues generated after the usage of the production equipment must be taken into account (Baykara, 2018).

Efficiency

The efficiency of a PEM electrolyser varies depending on the quality of the materials used, the design of the electrolyser, the operating temperature, the pressure, and the concentration of the electrolytes. In general, a typical Low Heating Value (LHV) efficiency of PEM electrolyser can range from 67% to 82%. To improve efficiency, efforts should focus on optimising the geometry of the electrolytic cell, using more efficient catalysts and optimising the operating conditions of the electrolyser (Wang et al., 2022).

Economic

The current lack of extensive green hydrogen production is mainly due to poor economic competitiveness. Therefore, most of the hydrogen production is done using fossil fuels. The major costs of green hydrogen production are related with the cost of renewable electricity, the efficiency of the electrolysis process and the cost of the electrolysis equipment (Yue et al., 2021).

Energy

In the case of green hydrogen, it is estimated that the production of 1 kg requires 50-55 kWh of electricity, which is considered high energy consumption compared to some fossil fuels (Kurrer, 2020). This consumption depends on the efficiency of the electrolysis process (Antweiler, 2020).

3 Process simulation

The PEM electrolyser flowsheet considered in this study is simulated using Aspen HYSYS software. The PEM model is implemented using Aspen Customer Module (ACM) software. The schematic used in Aspen HYSYS for the simulation of the entire system is shown in Figure 1, clearly depicting the division between the cathode (C) and anode (A) sides.

3.1 Simulation model

Voltage model

The PEM electrolyser's voltage model is determined by Equation (3), which calculates the total voltage required for a single cell to perform the electrolysis process. This model consists of several components, including the ideal voltage and the minimum voltage required, different loses, and factors resulting from the activation of the reaction incurred throughout the process (Colbertaldo et al., 2017; AspenTech, 2021).

$$V_{\text{cell}} = V_{\text{id.}} + \Delta V_{\text{act.}} + \Delta V_{\text{ohm.}} + \Delta V_{\text{diff.}} + \Delta V_{\text{par.}} \quad (3)$$

The minimum voltage required to initiate an electrolysis process in a cell is known as the ideal voltage ($V_{id.}$), as described by Equation (4).

$$V_{\text{id.}} = \frac{1}{\text{nF}}(\Delta G + RT_{\text{op.}} ln(\frac{p_{H_2} + p_{O_2}^{0.5}}{a_{H_2O}})) \quad (4)$$

Where n is the number of electrons, F is the Faraday's constant, ΔG is the Gibbs free energy value, R is the gas constant, $T_{op.}$ is the operational temperature in the cell, p is the partial pressure for both elements H_2 and O_2 and a_{H2O} is the water activity value. For reactions to take place, an activation voltage ($\Delta V_{act.}$), Equation (5), is required, based on the Tafel equation and incorporating Butler-Volmer's simplification (García-Valverde et al., 2012).

$$\Delta V_{\text{act.}} = \Delta V_{act,cat.} + \Delta V_{act,an.} \quad (5)$$

Where $\Delta V_{act,cat.}$ is the activation voltage in the cathode side and $\Delta V_{act,an.}$ is the anode side voltage activation. These activation voltages have the same equation on both sides described in Equation (6).

$$\Delta V_{act.,x} = \frac{\text{RT}_{\text{op.}}}{\alpha_x \text{nF}} \ln\left(\frac{i_u}{i_{0,x}}\right) \quad (6)$$

Where x represents the anode or the cathode, R is the gas constant, $T_{op.}$ is the operational temperature in the cell, α_x is the charger transfer coefficient, n is the number of electrons, F is the Faraday's constant, i_u is the useful current density and $i_{0,x}$ is the exchange current density, which depends on the temperature associated with the Butler-Volmer's Equation (García-Valverde et al., 2012). According to Ohm's law, the electrical losses ($\Delta V_{ohm.}$) occurring in anode, cathode and membrane during the electrolysis process are represented by Equation (7).

$$\Delta V_{\text{Ohm.}} = (R_{\text{cat.}} + R_{\text{an.}} + R_{\text{mem.}})\ i_u A_{\text{cell}} \quad (7)$$

Where $R_{cat.}$ is the cathode side resistance and is calculated using Equation (8), $R_{an.}$ is the anode side resistance and is calculated using Equation (8), $R_{mem.}$ is

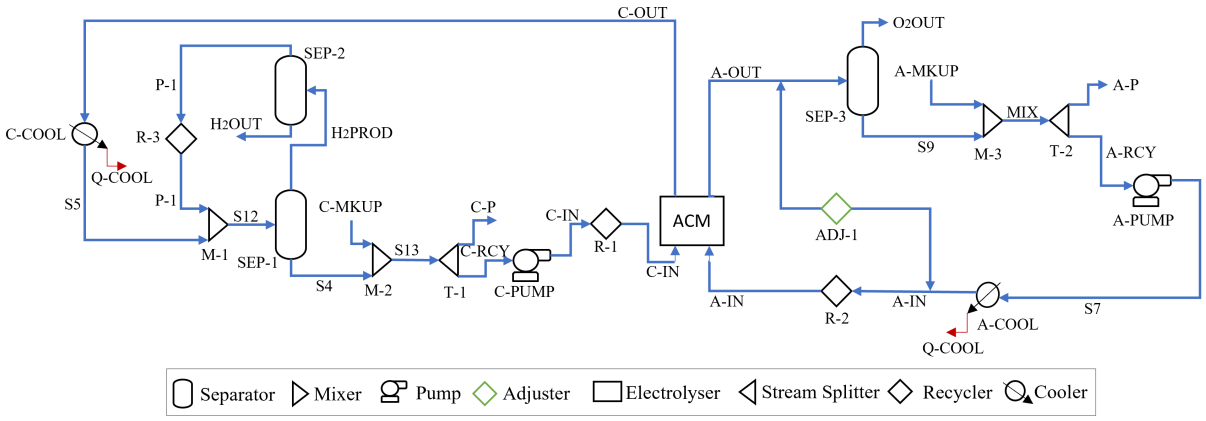


Figure 1. Process simulation schematic in Aspen HYSYS.

the membrane resistance and is calculated using Equation (9), i_u is the useful current density and A_{cell} is the active cell area.

$$R_x = \frac{t_x \rho_x}{A_x} \tag{8}$$

Where x represents the anode or the cathode, t_x is the electrode thickness, ρ_x is the resistivity and A_x is the active electrode area.

$$R_{mem.} = \frac{t_{mem}}{\sigma_{mem.} A_{mem.}} \tag{9}$$

Where $t_{mem.}$ is the membrane thickness, $\sigma_{mem.}$ is the conductivity based on the Springer model (Springer et al., 1991) and $A_{mem.}$ is the active membrane area. Diffusion voltage ($\Delta V_{diff.}$), Equation (10), represents the diffusion losses that occur when mass transport is hindered by the concentration gradient between the membrane surface and the main stream where the reaction takes place. These losses are the result of mass transport limitations due to the concentration gradient.

$$\Delta V_{diff.,x} = \frac{RT_{op.}}{\alpha_x nF} ln\left(\frac{i_L}{i_L - i_u}\right) \tag{10}$$

Where x represents the anode or the cathode, R is the gas constant, $T_{op.}$ is the operational temperature in the cell, α_x is the charger transfer coefficient, n is the number of electrons, F is the Faraday's constant, i_u is the useful current density and i_L is the limiting current density (assumed as 6 A/cm^2 the maximum current density). Parasitic losses (ΔVpar.), are typically expressed as a change in current rather than an increase in voltage. Essentially, the current efficiency is determined by the ratio of the input current to the useful current, Equation (11). This ratio is evaluated using the Faraday efficiency, which in the case of a PEM system, it is common to be close to 100%. Consequently, the Faraday efficiency used in the simulations is 99%.

$$\eta_{far.} = \frac{I_u}{I_{stack}} \tag{11}$$

Where I_u is the useful current calculated by multiplying the current density (i_u) by the active area of the cell (A_{cell}) and I_{stack} is the current in the cell.

Mass balance

The material balance evaluation in the electrolysis process is divided between the anode and cathode sides, and it is based on the assessment of the various flows involved. These flows include the water flow input, hydrogen production as described by Equation (12), oxygen production, electro-osmotic, diffusivity losses as described by Equations (13) and (14), respectively, and the pressure flow compensation as described by Equation (15).

$$\dot{N}_{H_2} = \frac{i_u A_{cell} N_{cells}}{nF} \tag{12}$$

Where i_u is the useful current density, A_{cell} is the active cell area, N_{cells} is the number of cells in the stack, n is the number of electrons and F is the Faraday's constant.

$$\dot{N}^{e-o}_{H_2O} = \frac{n_d i_u A_{cell} N_{cells}}{F} \tag{13}$$

Where n_d is the coefficient related with the humidification of the membrane extracted from (Colbertaldo et al., 2017), i_u is the useful current density, A_{cell} is the active cell area, N_{cells} is the number of cells in the stack and F is the Faraday's constant.

$$\dot{N}^{Diff.}_{H_2O} = \frac{D^{eff.}_{H_2O} \Delta C A_{cell} N_{cells}}{t_{mem}} \tag{14}$$

Where $D^{eff.}_{H_2O}$ is the diffusivity function based in (Aspen Technology, 2021), ΔC is the comparison water composition in the anode and cathode side, A_{cell} is the active cell area, N_{cells} is the number of cells in the stack and $t_{mem.}$ is the membrane thickness.

$$\dot{N}_{H_2O} = -\frac{K_{\text{Darcy}} A_{\text{cell}} \rho_{H_2O} \left(P_{\text{cat.}} - P_{\text{an.}}\right)}{\mu_{H_2O}} \quad (15)$$

Where K_{Darcy} is the membrane permeability, A_{cell} is the active cell area, ρ_{H2O} is the water density, $P_{cat.}$ and $P_{an.}$ are the pressure value in the cathode and anode side respectively and μ_{H2O} is the water viscosity.

Energy Balance

The energy balance is determined by comparing the energy inputs and outputs of the system equal to the total energy capacity. The inputs include the electrical power and the energy content of the inlet water flow, while the outputs encompass the heat losses (as described by Equation (16) as well as the outflow energy from both the anode and cathode sides.

$$Q_{\text{loss}} = h_{\text{free}} A_{\text{ext}} \left(T_{\text{op.}} - 10 - T_{\text{std.}}\right) \quad (16)$$

Where h_{free} is the heat transfer coefficient based in (AspenTech, 2021), $A_{ext.}$ is the exterior area (AspenTech, 2021), $T_{op.}$ is the operational temperature in the cell and $T_{std.}$ is the standard temperature.

3.2 *Process validation*

The simulated process is validated using simulated data, at an operating temperature of 55°C and an operating pressure of 30 bar, with the model presented in (Colbertaldo et al., 2017). The chosen operating conditions are based on the literature review performed during this study. Figure 2 illustrates the polarization curve demonstrating the relationship between the voltage cell and the current density. Furthermore, it provides insight into how various voltages incorporated in the model change as the current density increases. The specific comparison is made at a current density of 1,3 A/cm^2, where the voltage value for the simulated plant in this article is known to be 2,27 V. For the same data point in the reference article, the voltage is observed to fall between the values of 2,2 and 2,3 V. For the rest of the data points, the adjustment between the two models is carried out in a similar manner. Thus, a correlation can be drawn between the figure presented in this paper and the one found in the reference paper, validating the simulated model.

4 Sustainability analysis

The research on the application of the GREENSCOPE indicators to evaluate a PEM electrolysis process is scares. Following indicators are used to assess the sustainability of hydrogen production plants that utilize PEM technology.

4.1 *GREENSCOPE methodology*

Using the approach of sustainability and aiming to measure sustainability in any new or existing chemical

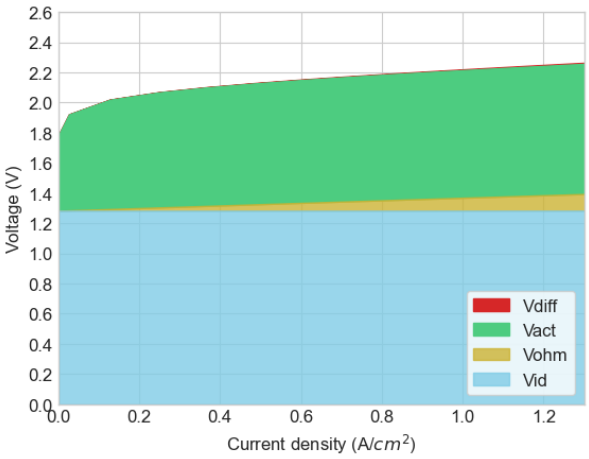


Figure 2. Polarization curve.

process throughout its life cycle analysis, the United States Environmental Protection Agency (US EPA) created the GREENSCOPE tool (EPA, 2015). The methodology in this tool is based on a set of metrics, GREENSCOPE indicators, used to evaluate the environmental performance of chemical products and processes in four different principal areas: Environmental, Efficiency, Economic and Energy. For the normalisation of the GREENSCOPE indicators, Equation (17) is used, which compares the actual process scenario with the best-case scenario of 100% sustainability and the worst-case scenario of 0% sustainability. The GREENSCOPE tables provide a comprehensive set of indicators and their corresponding parameters for calculating both the best and worst case scenarios. This Equation allows the comparison between different process (Li et al., 2016; Lima et al., 2016).

$$IndicatorScore = \frac{Actual - Worst}{Best - Worst} \; x \; 100 \; [\%] \quad (17)$$

4.2 *GREENSCOPE indicators*

To calculate the indicators selected for the PEM electrolyser, the operating point of the simulation is chosen corresponding to an electrical power of 6 kW, a temperature of 55 °C and a pressure of 30 bar. This particular operating point is chosen because it is of a similar magnitude to the operating point used in the AWE simulation available in literature, a more mature technology, is selected for the purpose of comparing the sustainability analyses (Sánchez et al., 2020; Hancke et al., 2022). The indicators selected for the sustainability analysis comparison of the simulated electrolyser are as follows: Global warming potential (GWP) – Environmental indicator, Mass Loss Index (MLI) – Efficiency indicator, Fractional Water Consumption (FWC) – Efficiency indicator, Specific Energy Costs ($C_{E,Spec.}$) – Economic indicator, Resource Energy Efficiency (η_E) – Energy indicator

(Ruiz-Mercado et al., 2014). Table 1 shows the data for the selected indicators. Note that for the $C_{E,Spec.}$ the energy and the product cost are calculated by using the model develop by (Jovan & Dolanc, 2020). This model requires estimation of the CAPEX of the simulated plants, that are calculated using the estimation model develop by (Reksten et al., 2022). For the other indicators, data is taken directly from the simulations.

Table 1. Data used for the GREENSCOPE indicators.

	PEM	AWE	Indicator
H_2 (kg/h)	0,101	0,220	GWP, MLI, FWC, η_E
O_2 (kg/h)	0,026	1,355	MLI
CO_2 (kg/h)	0,000	0,000	GWP
H_2O (kg/h)	0,002	0,002	FWC, η_E
$Prod._C$(\$/kg H_2)	7,170	6,249	$C_{E,Spec.}$
$En._C$(\$/kg H_2)	1,779	1,363	$C_{E,Spec.}$

5 Process scale-up

The linearisation method has been chosen to transform the data, making it suitable for the computation of regression models. This methodology is introduced to extrapolate the data, enabling a comparative analysis with commercially available electrolysers capable of generating greater quantities of hydrogen. The simulated data pertaining to the operating conditions of 30 bar pressure and 55°C temperature is used for scale-up purposes.

5.1 Regression and linearization model

The objective of data linearisation is to apply a regression model that initially do not have a linear dependence (James et al., 2021). In this study, simulated cell voltage cell (Vcell), specific work (Spc. work) and efficiency (η) data were taken and scaled-up as a function of current intensity. These parameters and ratios are typically the ones present in the reference article for the comparison of the simulated and scaled data of electrolysers present in the market (Buttler & Spliethoff, 2018). Table 2 presents the detailed explanation of the relationships of the variables for which regression models have been sought, including the data transformations performed and the variables to which it applies. The table also includes the regression models ultimately used, along with their corresponding R-squared values determining their suitability for use. Notably, all R-squared values are close to 1, indicating the high degree of fit and confirming the suitability of the generated regression models for the study's purposes. As an example, Figure 3 shows the application of a square root transformation to the abscissa results

Table 2. Data linearisation

X	Y	Reg. model	R-sq.
i_u	Vcell	$y = 0,39\ \sqrt{x} + 1,85$	0,96
i_u	Spc. work	$y = 0,42\ \log(x) + 5,27$	0,94
i_u	η	$y = -0,04\ \log(x) + 0,57$	0,99

in the linearisation of the data, which is then modelled using a regression model.

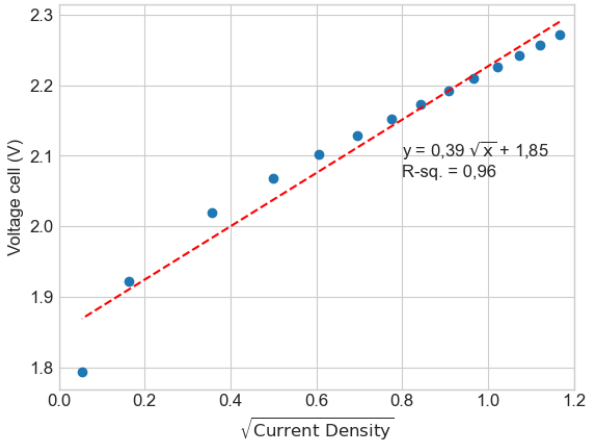


Figure 3. Plot of Vcell and Curr.Den linearisation

6 Results

The results of the Aspen HYSYS simulation, the sustainability analysis and the scaling-up process are presented below.

6.1 Simulation results

Figure 4 provides the correlation between voltage cell and current density, as well as the relationship between current density and efficiency. Where efficiency is defined as the ratio of energy extracted from the process in the form of hydrogen, using its LHV, and the amount of electrical energy input to the process. Solid lines are used to represent voltage cell, while dashed lines indicate the evolution of efficiency. For the various simulated points, the voltage of the cell is different for the same value of current densities. This is directly correlated with the amount of hydrogen produced. In other words, when less product is extracted, higher losses occur, resulting in a higher voltage for the cell. In terms of efficiency, a higher voltage increases hydrogen production and increases losses. Consequently, the simulation results represented by blue lines, corresponding to the highest pressure, has the worst performance.

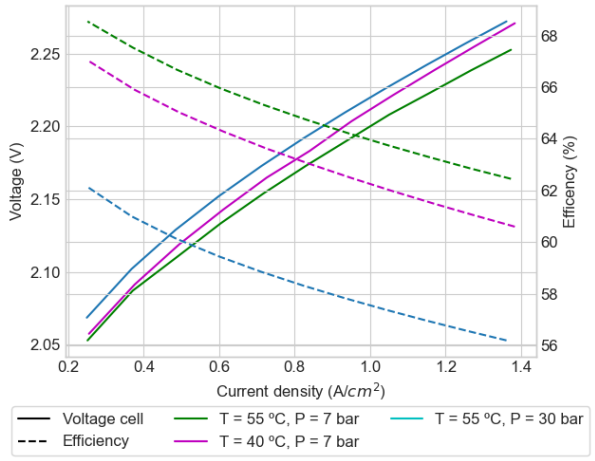


Figure 4. Plot of voltage cell and Efficiency of PEM electrolyser

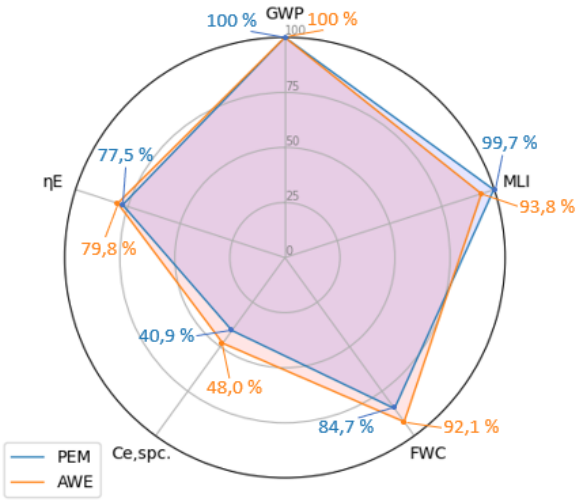


Figure 5. Plot of GREENSCOPE indicators for PEM and AWE

6.2 PEM sustainability

The GWP indicator is employed as a precise and visually effective representation of the complete absence of gas or pollutant emissions in the production of green hydrogen when the energy used is from a renewable source. The score of this indicator is 100 % for both production methods, that evidence the zero emission of CO_2 pollutants during the process. The results of the MLI indicator shows that the PEM process achieves a normalization value of 99,74 %, whereas the AWE process yields a value of 93,84 %. This indicates that AWE processes are less efficient in terms of hydrogen production. In terms of the FWC indicator, the AWE process demonstrates superior efficiency in utilizing the required water resource for its operation compared to other processes, such as PEM. While the PEM normalization of the indicator reflects a level of 84,68 %, the AWE process achieves significantly higher levels, reaching close to 92,00 %. The normalized $C_{E,spec.}$ indicator value for the PEM process is 40,91 %, while for the AWE process it is 48,04 %, indicating that energy cost has a greater impact on both processes. It is possible to see how the more mature AWE technology has a better cost distribution, although the difference is not very large. The ηE indicators for hydrogen production using PEM technology and AWE are 77,50 % and 79,84 % respectively, indicating that AWE has slightly higher efficiency compared to PEM. Figure 5 displays the normalized values for all the indicators, along with the comparison between PEM and AWE technologies, depicted in blue and orange respectively. The variation in both electrolysers can be justified by method of operation and technology differences.

6.3 Scale-up analysis

To verify the suitability of the simulated installation for large-scale hydrogen production, the benchmarking study conducted by Buttler and Spliethoff is used as a referenced. This study includes graphical representations of the current market's PEM electrolysers (Buttler & Spliethoff, 2018). The figure 6 displays the current density against the voltage, specific work, and cell efficiency. The polarization line is shown in blue, while the efficiency using the lower heating value (LHV) as the reference value is displayed in grey dashed lines. Furthermore, a second y-axis is added, representing the specific work values. The minimum voltage is approximately 1,8 V. In commercial electrolysers, this value consistently remains below 1,75 V and approaches 1,5 V. This discrepancy arises because the results do not converge at lower power levels due to small scale nature of the simulated process. Nevertheless, the results for power levels of 1 kW and above are satisfactory and facilitated a comprehensive analysis. The specific work values adequately match those shown in the reference article. Therefore, it can be concluded that this variable can be compared with that of real electrolyzers. Additionally, efficiency values have been obtained that are realistic and, when compared to those shown in the article, indicate that the scaling of the simulation is satisfactory for the efficiency parameter. In summary, the initial attempt to scale-up the process has yielded favorable results. However, it must be acknowledged that certain challenges, exemplified by the encountered setback related to lower values, have surfaced. To further enhance these endeavors aimed at cost-effective design improvements, consideration should be given to exploring alternative scaling methodologies. One potential avenue involves

the utilization of techniques such as Buckingham's π theorem (Polverino et al., 2019) or the incorporation of Artificial Neural Networks (ANN) strategies (Tian, 2020), both have demonstrated effectiveness in the context of enlarging hydrogen fuel cells. These approaches hold promise for facilitating significant advancements.

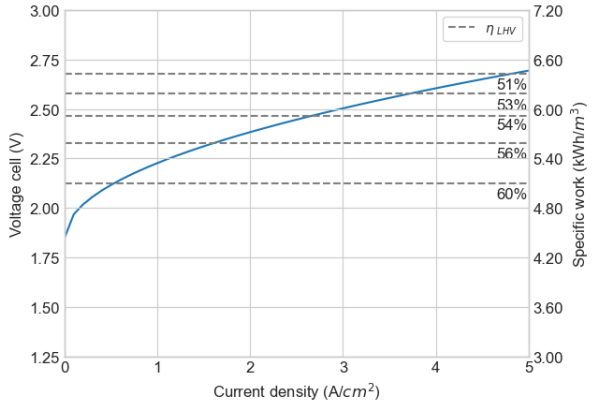


Figure 6. Plot of Scaled-up process variables of PEM electrolyser

7 Conclusions

The PEM simulation exhibited satisfactory performance when compared to the reference simulation. Specifically, the operating range of 1-6 kW was simulated successfully for subsequent analysis. The analysis shows that the higher temperature and electrical power levels increased hydrogen production, while pressure had an inversely proportional effect. This finding is consistent with observations made in various real and simulated PEM electrolysers documented in the literature. The sustainability analysis performed consisted of calculating indicator values according to the GREENSCOPE methodology, which was used for the first time in this paper to evaluate hydrogen production methods. In addition, a sustainability comparison was made between PEM and AWE technologies. The environmental indicator, GWP, was found to be 100 for both technologies as green hydrogen production with renewable energy sources does not generate CO_2 emissions. With regards to the efficiency indicators, the proportion of hydrogen produced, MLI, was found to be higher for PEM technology than for AWE, while water consumption, FWC, was better for AWE technology. These variations can be justified by technology differences present in both electrolysers. In terms of the economic indicator, $C_{E,spec.}$, it was observed that the weight of energy costs was higher in the case of PEM technology. Finally, the efficiency indicator showed that the energy efficiency, ηE, was slightly worse for PEM technology. The scaling-up process employed data linearisation and regression techniques. Through this approach, the simulation demonstrated satisfactory comparability with commercially available PEM electrolysers. While scaling-up processes for hydrogen fuel cells simulations using PEM technology are documented, methods such as Artificial Neural Networks (ANN) or Buckingham π theorem for electrolysers simulations are yet to be explored. The application of such methods holds promise for significant cost reductions in the production of commercialised electrolysers, further enhancing viability.

Acknowledgement

The authors acknowledge the Department of Electrical Engineering, Information Technology and Cybernetics (EIK) at University of South-Eastern Norway (USN) for facilitating with computer resources and funding support. The authors would also like to thank ERASMUS+ exchange program that allows for students to study at universities in European Union.

References

Antweiler, W. (2020). *What role does hydrogen have in the future of electric mobility?* Retrieved 2023-02-27, from `https://wernerantweiler.ca/blog.php?item=2020-09-28`

AspenTech. (2021). Water Electrolysis with PEM.

Baykara, S. Z. (2018, June). Hydrogen: A brief overview on its sources, production and environmental impact. *International Journal of Hydrogen Energy*, *43*(23), 10605–10614. doi: 10.1016/j.ijhydene.2018.02.022

Buttler, A., & Spliethoff, H. (2018, February). Current status of water electrolysis for energy storage, grid balancing and sector coupling via power-to-gas and power-to-liquids. *Renewable and Sustainable Energy Reviews*, *82*, 2440–2454. doi: 10.1016/j.rser.2017.09.003

Colbertaldo, P., Gómez Aláez, S. L., & Campanari, S. (2017, December). Zero-dimensional dynamic modeling of PEM electrolyzers. *Energy Procedia*, *142*, 1468–1473. doi: 10.1016/j.egypro.2017.12.594

EPA, U. (2015, December). *Assessing Chemical Process Sustainability with GREENSCOPE* [Data and Tools]. Retrieved from `https://www.epa.gov/chemical-research/assessing-chemical-process-sustainability-greenscope`

García-Valverde, R., Espinosa, N., & Urbina, A. (2012, January). Simple PEM water electrolyser model and experimental validation. *International Journal of Hydrogen Energy*, *37*, 1927–1938. doi: 10.1016/j.ijhydene.2011.09.027

Hancke, R., Holm, T., & Ulleberg, O. (2022, June). The case for high-pressure PEM water electrolysis. *Energy Conversion and Management*, *261*, 115642. doi: 10.1016/j.enconman.2022.115642

IRENA. (2021). *Hydrogen.* Retrieved from `https://www.irena.org/Energy-Transition/Technology/Hydrogen`

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning with applications in R* (Vol. 6). Retrieved from `https://www.tandfonline.com/doi/full/10.1080/24754269.2021.1980261`

Jovan, D. J., & Dolanc, G. (2020, January). Can Green Hydrogen Production Be Economically Viable under Current Market Conditions. *Energies*, *13*(24). doi: 10.3390/en13246599

Li, S., Mirlekar, G., Ruiz-Mercado, G. J., & Lima, F. V. (2016, September). Development of chemical process design and control for sustainability. *Processes*, *4*(3), 23. (Number: 3 Publisher: Multidisciplinary Digital Publishing Institute) doi: 10.3390/pr4030023

Lima, F., Li, S., Mirlekar, G., Sridhar, L., & Ruiz-Mercado, G. (2016). Chapter five - modeling and advanced control for sustainable process systems. In G. Ruiz-Mercado & H. Cabezas (Eds.), *Sustainability in the design, synthesis and analysis of chemical engineering processes* (p. 115-139). Oxford: Butterworth-Heinemann. doi: https://doi.org/10.1016/B978-0-12-802032-6.00005-0

Noussan, M., Raimondi, P. P., Scita, R., & Hafner, M. (2021, January). The Role of Green and Blue Hydrogen in the Energy Transition—A Technological and Geopolitical Perspective. *Sustainability*, *13*. doi: 10.3390/su13010298

Polverino, P., Bove, G., Sorrentino, M., Pianese, C., & Beretta, D. (2019, September). Advancements on scaling-up simulation of Proton Exchange Membrane Fuel Cells impedance through Buckingham Pi theorem. *Applied Energy*, *249*, 245–252. doi: 10.1016/j.apenergy.2019.04.067

Reksten, A. H., Thomassen, M. S., Møller-Holst, S., & Sundseth, K. (2022, November). Projecting the future cost of PEM and alkaline water electrolysers; a CAPEX model including electrolyser plant size and technology development. *International Journal of Hydrogen Energy*, *47*(90), 38106–38113. doi: 10.1016/j.ijhydene.2022.08.306

Ruiz-Mercado, G. J., Gonzalez, M. A., & Smith, R. L. (2014, April). Expanding GREENSCOPE beyond the gate: a green chemistry and life cycle perspective. *Clean Technologies and Environmental Policy*, *16*(4), 703–717. doi: 10.1007/s10098-012-0533-y

Springer, T. E., Zawodzinski, T. A., & Gottesfeld, S. (1991, August). Polymer Electrolyte Fuel Cell Model. *Journal of The Electrochemical Society*, *138*(8), 2334. doi: 10.1149/1.2085971

Sánchez, M., Amores, E., Abad, D., Rodríguez, L., & Clemente-Jul, C. (2020, February). Aspen Plus model of an alkaline electrolysis system for hydrogen production. *International Journal of Hydrogen Energy*, *45*(7), 3916–3929. doi: 10.1016/j.ijhydene.2019.12.027

Tian, P. (2020). *Performance prediction of PEM fuel cell using artificial neural network machine learning* (Doctoral dissertation, UC Irvine). Retrieved from `https://escholarship.org/uc/item/47t0j1pg`

Wang, T., Cao, X., & Jiao, L. (2022, June). PEM water electrolysis for hydrogen production: fundamentals, advances, and prospects. *Carbon Neutrality*, *1*(1), 21. doi: 10.1007/s43979-022-00022-8

Younas, M., Shafique, S., Hafeez, A., Javed, F., & Rehman, F. (2022, May). An Overview of Hydrogen Production: Current Status, Potential, and Challenges. *Fuel*, *316*, 123317. doi: 10.1016/j.fuel.2022.123317

Yue, M., Lambert, H., Pahon, E., Roche, R., Jemei, S., & Hissel, D. (2021, August). Hydrogen energy systems: A critical review of technologies, applications, trends and challenges. *Renewable and Sustainable Energy Reviews*, *146*, 111180. doi: 10.1016/j.rser.2021.111180

Process Simulation and Cost Optimization of CO_2 Capture Configurations in Aspen HYSYS

Lars Erik Øi, Madhawee Anuththara, Shahin Haji Kermani, Mostafa Mirzapour, Soudeh Shamsiri and Sumudu Karunarathne

Department of Process, Energy and Environmental Technology, University of South-Eastern Norway

Lars.oi@usn.no

Abstract

A CO_2 capture process from a natural gas based power plant has been simulated and cost estimated using an equilibrium-based model in Aspen HYSYS using the amine acid gas package. The aim has been to calculate cost optimum process parameters for the standard process and also for a vapor recompression process. After process simulation using Aspen HYSYS, the process equipment was dimensioned and cost estimated using Aspen In-plant. The Enhanced Detailed Factor (EDF) method was used to select factors to calculate the total investment. Operating cost for heat and electricity was calculated from the simulation with estimated cost on consumed heat and electricity. The cost was calculated to 21.2 EURO per ton CO_2 removed and a vapor recompression process was calculated to 21.6 EURO per ton. A recompression case with 1.2 bar flash pressure was calculated to 21.3 EURO/ton CO_2. The ΔT_{MIN} in the amine/amine heat exchanger was varied, and the optimum at 15°C was 20.9 EURO per ton CO_2. The vapor recompression alternative was in this work slightly more expensive than the traditional case. In earlier works, the vapor recompression process has been claimed to be more economical than the standard process. The difference in this work is mainly due to different cost estimates of the compressor investment. This work shows that Aspen HYSYS is well suited for optimizing process parameters in a CO_2 capture process with and without vapor recompression.

Keywords: *Carbon capture, Aspen HYSYS, simulation, cost estimation, optimization*

1. Introduction

CO_2 capture based on absorption into an amine followed by desorption is an established method to reduce CO_2 emissions. Much work has been performed on simulation and cost estimation of CO_2 capture processes, especially from natural gas based power plants. A traditional tool has been an equilibrium-based model in Aspen HYSYS using the amine acid gas package. The aim has often been to calculate cost optimum process parameters for a standard process. In this work, the main aim has been to calculate cost optimum process parameters for a standard CO_2 capture process. A special aim has been to compare the standard process with a process based on vapor recompression. It shows that it is difficult to state whether the vapor recompression process is more economical than a standard CO_2 capture process.

2. Literature, Process Description and Specifications

2.1 Literature

There are several papers presenting results from process simulation and cost estimation of CO_2 capture plants (Manzolini et al., 2015; Luo and Wang, 2016; Nwaoha et al., 2018; Hasan et al., 2021). This work is a continuation of previous work at the Telemark University College and the University of South-Eastern Norway (USN). Some references are (Kallevik, 2010; Øi, 2012; Aromada and Øi, 2017; Øi et al., 2020; Øi et al., 2021; Shirdel et al. (2022). These projects have involved process simulation, dimensioning and cost estimation of CO_2 capture using the process simulation tool Aspen HYSYS. Capture rate, energy demand and capture cost per ton CO_2 have been calculated. Many of the projects have optimized parameters by changing one process parameter at a time, such as the minimum temperature difference in the main heat exchanger.

In the literature there have been presented many suggestions for process improvements using different process configurations (Cousins et al., 2011; Moullec et al., 2011; Dubois and Thomas, 2017). A simple alternative is vapor recompression where regenerated amine is depressurized into a flash tank, and the flash gas is recompressed and sent to the bottom of the desorber. Cost optimization of vapor recompression has been perfomed by Fernandez et al. (2012), Øi et al. (2014), Aromada and Øi (2017), Øi et al. (2017) and Øi

et al. (2021). This work is based on a Master group project (Kermani et al., 2022). In addition to the project work, simulation and cost estimation of the vapor compression process from 1.2 to 2 bar was also included.

2.1. Process description of a standard process

Fig. 1 shows a typical process for CO_2 capture using an amine absorbent. In this method, CO_2 is absorbed and captured in an aqueous amine solution, in which flue gas is passed through. The CO_2-rich amine is then sent to a stripper, is heated with steam, and as a result CO_2 is released from the solution. In the figure, a gas cooler before the absorber and a water wash are shown, but these units are not simulated in this work.

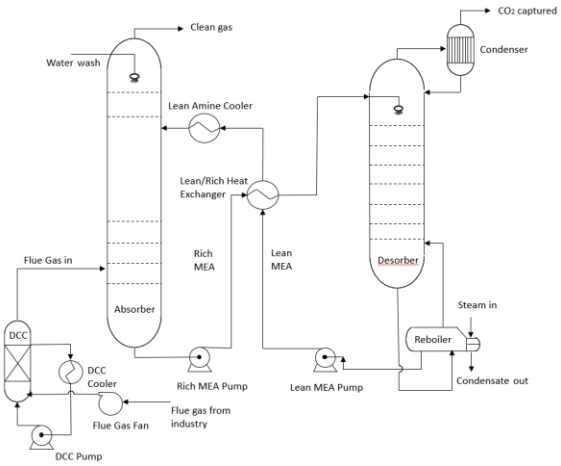


Figure 1: Process flow diagram of a standard amine-based CO_2 capture process (Aromada et al., 2020)

3 Specifications and simulations

3.1 Specifications and simulation of standard CO_2 capture process

The specifications for the base case is given in Table 1. The calculation sequence is similar to earlier works (Aromada and Øi, 2015; Øi et al., 2020; Øi et al, 2021). The absorption column is calculated first based on the inlet gas and the estimated lean amine flow (which is specified in the first iteration). The amine from the bottom of the absorption column is sent to regeneration via the rich/lean heat exchanger. The amine flow is entering the desorption column which separates the feed into CO_2 product at the top and hot regenerated amine at the bottom. The regenerated amine is returned via the lean/rich heat exchanger and the lean cooler to the recycle block. Due to water loss in the process, water must be added to the process. The make-up water was adjusted manually. The specifications in Table 1 aim at a 90 % CO_2 removal efficiency and gives the result of 7.7 °C in the lean/rich heat exchanger. The simulations were performed in Aspen Plus V12.

Table 1. Aspen HYSYS model parameters and specifications for the base case alternative

Parameter	
Inlet flue gas temperature [°C]	40.0
Inlet flue gas pressure [kPa]	110
Inlet flue gas flow rate [kmol/h]	85000
CO_2 content in inlet gas [mole %]	3.73
Water content in inlet gas [mole %]	6.71
Lean amine temperature [°C]	40.0
Lean amine pressure [kPa]	110.0
Lean amine rate [kg/h]	110000
MEA content in lean amine [mol-%]	11.21
CO_2 content in lean amine [mol-%]	2.93
Number of stages in absorber [-]	10
Murphree efficiency in absorber [m^{-1}]	0.25
Rich amine pump pressure [kPa]	200.0
Rich amine temp. out of HEX [°C]	104.9
Number of stages in desorber [-]	6
Murphree efficiency in desorber [m^{-1}]	1
Reflux ratio in stripper [-]	0.3
Reboiler temperature [°C]	120.0
Lean amine pump pressure [kPa]	500.0

3.2 Specification of vapor recompression process

The Aspen HYSYS flowsheet for the base case is presented in Fig. 2. The flowsheet for the vapor recompression process is presented in Fig. 3. After the desorber, the amine is pressure reduced through a valve to a flash tank. The gas after the flash tank with atmospheric pressure (or higher) is compressed and sent back to the desorber. Except for this, the process is the same as in the base case.

3.3 Parameter variations

With a 110000 kg/h amine flowrate, absorption 10 stages, 90 % removal efficiency and 7.7 °C minimum approach temperature were obtained in the base case simulation. The minimum approach temperature was varied. For the vapor recompression case, the flash pressure was varied. In the parameter variation simulations, all other specified parameters were kept constant.

A possibility is to make use of the Case study function in Aspen HYSYS. In that case a series of calculations can be performed automatically keeping all other specified parameters constant.

3.4 Simulation and cost estimation procedure

The objective of this part is the estimation of the plant's total cost for the designed CO_2 capture process. Calculations are based on dimensions obtained from the simulation in Aspen HYSYS V12. A short version of the cost estimation procedure is as follows, similar to the procedure in Øi et al. (2020) and Øi et al. (2021):

- Calculation of each equipment cost using Aspen In-Plant Cost Estimator, based on equipment dimensioning parameters for the base case.
- Calculation of the total installation cost by applying the Enhanced Detailed Factor (EDF) method.
- Correction of total installation cost by the cost inflation index (conversion by year).
- Calculation of annualized capital expenditure (CAPEX) according to the discount rate and lifetime
- Calculation of annual operational expenditure (OPEX)
- Calculation of the total CO_2 capture cost based on the plant lifetime

3.5 Dimensioning for cost estimation

The estimation of packing height is based on a constant stage (Murphree) efficiency corresponding to 1 meter of packing. Murphree efficiencies were specified to 0.25 for the absorber and 1.0 for the desorber. Structured packing was assumed.

The estimation of absorption column diameter was based on a gas velocity of 2.5 m/s and for the desorption column a gas velocity of 1 m/s was assumed as in Øi et al. (2020) and Øi et al. (2021). The total height of the absorption column and desorption column were specified to be 25 m and 16 m respectively. The extra height is due to distributors, water wash packing, demister, gas inlet, outlet and sump. The pumps and the vapor compressor were specified to have 75 % adiabatic efficiency.

Overall heat transfer coefficient values were specified for the lean/rich heat exchanger 500 W/(m^2K), lean amine cooler 800 W/(m^2K), reboiler 1200 W/(m^2K) and condenser 1000 W/(m^2K). These values are the same as in Øi et al. (2021) except for the lean/rich heat exchanger number (changed from 550 W/m^2K), and slightly less than the numbers in Øi et al. (2020).

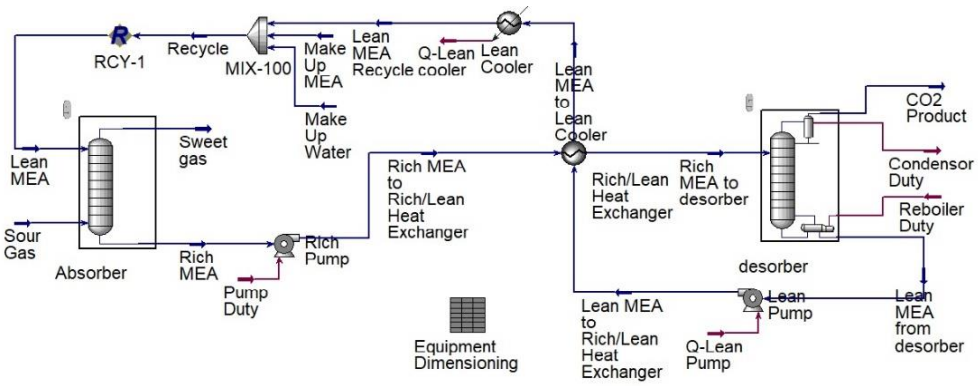


Figure 2. Aspen HYSYS flow-sheet of the base case simulation (from Kermani et al., 2022)

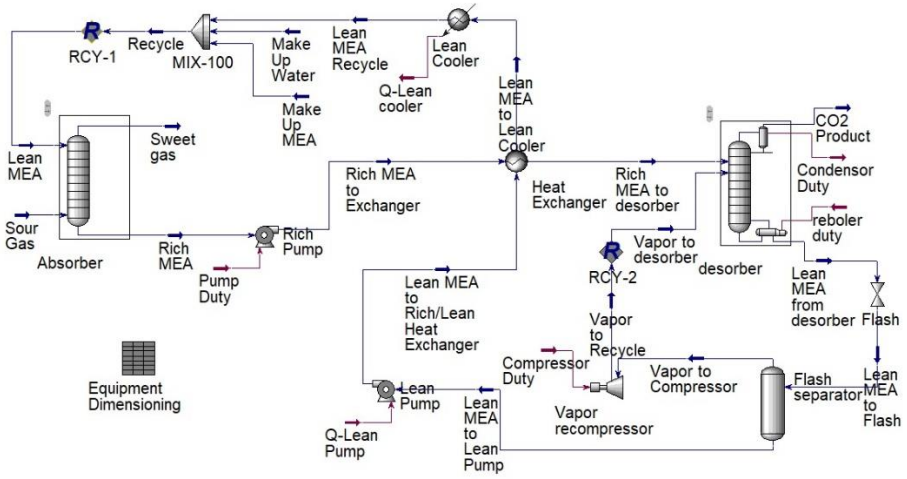


Figure 3. Aspen HYSYS flow-sheet of the vapour recompression case simulation (from Kermani et al., 2022)

3.6 Capital cost estimation methods

Equipment costs were calculated in Aspen In-plant Cost Estimator (version 12), which gives the cost in Euro (€) for Year 2016 (1st Quarter). Stainless steel (SS316) with a material factor of 1.75 was assumed for all equipment units, except for pumps and the vapor compressor where a material factor of 1.3 was used as in Øi et al. (2020) and Øi et al (2021).
In the EDF detailed factor method, each equipment cost in carbon steel was multiplied with an installation factor to obtain installed cost. 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. Installation factors from Aromada et al. (2021) were used.

Table 3. Cost calculation specifications

Parameter	**Value**
Plant lifetime	10 and 20 years
Discount rate	7.5 %
Maintenance cost	4 % of installed cost
Electricity price	0.06 EURO/kWh
Steam price	0.015 EURO/kWh
Annual operational time	8000 hours
Location	Rotterdam

3.7 Operating cost calculation

This project includes OPEX estimations for the use of electricity and steam (maintenance cost is not included). Operating cost specifications are given in Table 3. Electricity cost was specified to be 0.06 EURO/kWh (approximately 0.6 NOK/kWh). The steam cost was specified to be 25 % of the electricity cost, 0.015 EURO/kWh. This is reasonable for a case where the heat could be converted to electricity with 25 % efficiency. The detailed cost estimation of CAPEX, OPEX and NPV (net present value) were calculated in an internal spreadsheet in Aspen HYSYS.

4 Results and Discussion

4.1 Base case cost results

In Fig. 4, the results for the capital cost estimation of the base case are shown for all the equipment units. The total cost was calculated to 74.6 mill. EURO. The total cost per ton CO_2 removed was calculated to 21.2 EURO/ton CO_2. The numbers are low compared to many other estimations, but the values in Øi et al. (2020) are similar. One reason is that some equipment like pre-treatment and water wash is not included in these calculations. However, for optimization calculations, only the units in the recirculation are necessary to obtain a reasonable optimization.

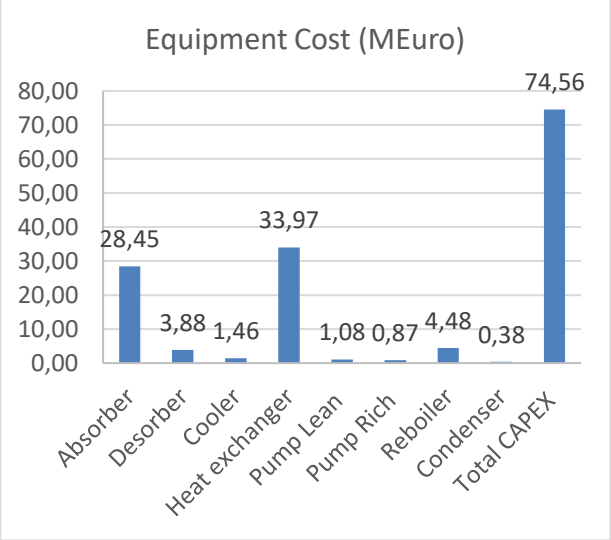


Figure 4: Total CAPEX and the cost of each piece of equipment for the base case (Kermani et al., 2022).

The equipment cost shows that the most expensive equipment units are the absorber and the main heat exchanger. This is traditional. Normally the absorber is the most expensive unit, so there is a possibility that the absorber cost is underestimated. The estimated column efficiency is 0.25 per meter packing height, which is optimistic compared to 0.15 in Øi et al. (2021). A water wash is normally a part of the absorber, and this cost is neglected in this work. The total cost is probably also underestimated because there are probably equipment unit details that are more complex than assumed. The operating cost is probably underestimated because the maintenance cost is not included. If both CAPEX and OPEX is underestimated to the same degree, the trade-off between them will give reasonable cost optimum parameters.

4.2 Vapour recompression case

The vapor recompression cost was calculated to 21.6 EURO/ton CO_2 for a flash pressure of 1 bar. This is slightly higher than the standard process,

and in this work this was not optimum. The cost was also calculated for a flash pressure of 1.2 bar, and the result was 21.3 EURO/ton CO_2. This was the optimum vapour recompression case, but it was still not optimum compared to the base case. In earlier work (Karimi et al., 2011; Øi et al., 2014), the vapour recompression case was estimated to be the most optimum process. The difference in this work is mainly due to different estimates of the compressor investment. It is possible to reduce the cost of the vapor recompression by optimizing the flash pressure as in Fernandez et al. (2012). In Øi et al. (2021), a flash pressure of 1.5 bar was the optimum in the vapour recompression case, but was not better economically than the standard process.

4.3 Optimum minimum temperature approach

The total cost was calculated for different temperature approaches. The result is shown in Fig. 5 with the base case shown for 7.7 K. The absorber packing height was 15 m in these optimizations. The optimum value was found as the one with minimum total cost at 15 K with 20.9 EURO/ton. Øi et al. (2014) and Aromada and Øi (2017) get about the same optimum. Values for the optimum minimum temperature approach in literature are often between 10 and 15 K.

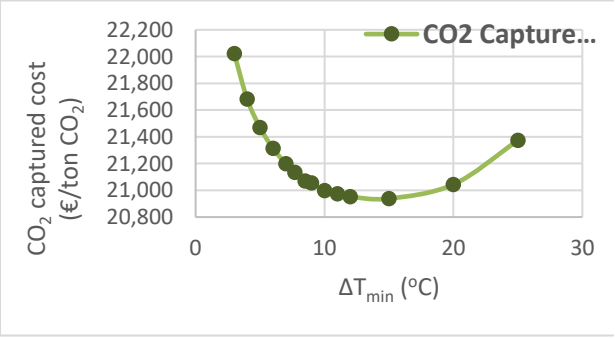


Figure 5. Optimization of minimum approach temperature for the base case (Kermani et al., 2022)

4.4 Comparison with earlier studies

The numbers in Table 4 show different literature sources with typical or optimized values for CO_2 capture rate, inlet CO_2 concentration, ΔT_{min}, absorber packing height and reboiler duty. The table shows that the calculated and estimated temperature approach, reboiler duty and absorber height are similar to values found in literature.

Table 4. Comparison of this study with previous base case scenarios

Study	CO_2 capture rate [%]	CO_2 con. [mol %]	ΔTmin [°C]	Absorber packing height [m]	Reboiler du [kJ/capture CO_2 kg]
Present work (base case)	90	3.73	7.7	10 (25-total height)	3757
Present work (ΔT_{min} =10)	90	3.73	10	10	3852
Aromada et al. [11]	85	3.73	10	10	3600
Øi et al. [15]	85	3.75	10	10	3650
Alhajaj et al. [24]	90	5	20	34.3	4484
Amrollahi et al. [25]	90	3.8	8.5	13	3740
Sipöcz et al. [26]	90	4.2	10	26.9 (Total height)	3930
Karimi et al. [9]	90	11.86	5	7	3545
	90	11.86	10	7	3611

Some of the numbers are optimized and some of them are typical or reasonable values. Most of them are for CO_2 capture processes for natural gas based power plants with about 4 mol-% CO_2 in the exhaust gas as in this work. The rounded values in the values shown in Table 4 indicate that there is need for further work to find optimum values for these parameters.

5. Conclusion

A CO_2 capture process from a natural gas based power plant has been simulated and cost estimated using an equilibrium-based model in Aspen HYSYS using the amine acid gas package. The aim has been to calculate cost optimum process parameters for the standard process and also for a vapor recompression process.

After process simulation using Aspen HYSYS, the process equipment was dimensioned and cost estimated using Aspen In-plant. The Enhanced Detailed Factor (EDF) method was used to select factors to calculate the total investment. Operating cost for heat and electricity was calculated from the simulation with estimated cost on consumed heat and electricity. The cost was calculated to 21.2 EURO per ton CO_2 removed and a vapor recompression process was calculated to 21.6 EURO per ton.

The ΔT_{MIN} in the amine/amine heat exchanger was varied, and the optimum at 15°C was 20.9 EURO per ton CO_2. The vapor recompression alternative also calculated with 1.2 bar flash pressure, was in this work slightly more expensive than the traditional case. In earlier works, the vapor recompression process has been claimed to be more economical than the standard process. The difference in this work is mainly due to different cost estimates of the compressor investment.

This work shows that Aspen HYSYS is well suited for optimizing process parameters in a CO_2 capture process with and without vapor recompression.

References

Ali, H. (2019) *Techno-economic analysis of CO_2 capture concepts.* PhD Thesis, University of South-Eastern Norway.

Alhajaj, A. (2016) 'A techno-economic analysis of post-combustion CO_2 capture and compression applied to a combined cycle gas turbine: Part II. Identifying the cost-optimal control and design variables', *International Journal of Greenhouse Gas Control*, 52:331-343, doi: 10.1016/j.ijggc.2016.07.008.

Amrollahi, P. A. M. et al. (2012) 'Optimized process configurations of post-combustion CO_2 capture for natural-gas-fired power plant – Power plant efficiency analysis', *International Journal of Greenhouse Gas Control*, 8:1-11, doi: 10.1016/j.ijggc.2012.01.005.

Aromada, S. A. and Øi, L.E. (2017) 'Energy and economic analysis of improved absorption configurations for CO_2 capture.' *Energy Procedia*, 114: pp. 1342-1351.

Aromada, S. A. et al. (2021) 'Capital cost estimation of CO_2 capture plant using Enhanced Detailed Factor (EDF) method: Installation factors and plant construction characteristic factors', *Int. J. Greenh. Gas Control*, vol. 110, no. 1, p. 103394, doi: 10.1016/j.ijggc.2021.103394.

Cousins A. et al. (2011) 'A survey of process flow sheet modifications for energy efficient CO_2 capture from flue gases using chemical absorption'. *International Journal of Greenhouse Gas Control*, 5(4):605-619.

Dubois, L. and Thomas, D. (2017) 'Comparison of various configurations of the absorption-regeneration process using different solvents for the post-combustion CO_2 capture applied to cement plant flue gases.' *International Journal of Greenhouse Gas Control*, 69:20-35.

Eldrup, N. (2016) *Installation factor sheet - Project Management and Cost Engineering*. Master's Course. University College of South-Eastern Norway, Porsgrunn.

Eldrup, N. et al. (2019) 'A Cost Estimation Tool for CO2 Capture Technologies', SSRN Scholarly paper ID 3366036, doi: 10.2139/ssrn.3366036

Fernandez, E. S. et al. (2012) 'Optimisation of lean vapour compression (LVC) as an option for post-combustion CO_2 capture: Net present value maximisation.' *International Journal of Greenhouse Gas Control*, 11:114-121.

GPSA (1987), *Engineering Data Book, 10. ed., Tulsa, Oklahoma: Gas Processing Suppliers Association, 1987.*

Hasan, S. et al. (2021) 'Improving the Carbon Capture Efficiency for Gas Power Plants through Amine-Based Absorbents', *Sustainability*, 13(1), doi: 10.3390/su13010072

Kallevik, O. B. (2010) *Cost estimation of CO_2 removal in HYSYS*. Master's Thesis, Telemark University College, Porsgrunn.

Karimi, M. et al. (2011) 'Capital costs and energy considerations of different alternative stripper configurations for post combustion CO_2 capture.' *Chemical Engineering Research and Design*, 89(8):1229-1236.

Kermani, S. H. et al. (2022) 'Process simulation and cost estimation of CO_2 capture configurations in Aspen HYSYS.' Master project, University of South-Eastern Norway.

Luo, X and Wang, M. (2016) 'Optimal operation of MEA-based post-combustion carbon capture for natural gas combined cycle power plants under different market conditions," *International Journal of Greenhouse Gas Control*, 48, pp. 312-320. doi: 10.1016/j.ijggc.2015.11.014.

Manzolini, G. et al. (2015) 'Economic assessment of novel amine based CO_2 capture technologies integrated in power plants based on European Benchmarking Task Force methodology'. *Applied Energy*, 138, pp. 546-548

Nwaoha, C. *et al.* (2018), '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 pp. 400-412.

Sipöcz, N. et al. (2011), 'Integrated modelling and simulation of a 400 MW NGCC power plant with CO_2 capture', *Energy Procedia*, 114: pp. 1342-1351.

Shirdel, S. *et al.* 'Sensitivity Analysis and Cost Estimation of a CO_2 Capture Plant in Aspen HYSYS'. *Chem. Engineering*, 6(2), 28, 2022.

Souders, M. and Brown, G. G. (1934) 'Design of Fractionating Columns I. Entrainment and Capacity.' *Industrial & Engineering Chemistry*, 26:98-103.

Øi, L. E. (2012) *Removal of CO_2 from exhaust gas*. PhD Thesis, Telemark University College, Porsgrunn.

Øi, L E. et al. (2014) 'Optimization of Configurations for Amine based CO_2 Absorption Using Aspen HYSYS.' *Energy Procedia*, 51: 224-233.

Øi, L. E. and Sundbø, E. (2017) '*Simulation and Economic Optimization of Vapor Recompression Configuration for Partial CO_2 capture*.' In the 58th Conference on Simulation and Modelling, SIMS 58, Reykjavik, Iceland.

Øi, L. E. *et al. (2021)* 'Automated Cost Optimization of CO_2 Capture Using Aspen HYSYS', Linköping Electronic Conference Proceedings SIMS 62, pp. 293-300. doi: 10.3384/ecp21185293

Performance assessment of a photovoltaic/thermal-powered absorption chiller for a restaurant

Nima Monghasemi [a,*], Stavros Vouros [b], Konstantinos Kyprianidis [c], Amir Vadiee [d]

[a,c] *School of Business, Society, and Engineering, Division of Sustainable Energy Systems, Mälardalen University, Sweden*

[b] *Academy for Innovation, Design and Technology, Department of Product Realization, Mälardalen University, Sweden*

[d] *School of Business, Society, and Engineering, Division of Sustainable Environment and Construction, Mälardalen University, Sweden*

nima.monghasemi@mdu.se

Abstract

In recent years the demand for cooling in buildings has grown steadily due to factors such as climate change and increased use of technology in Sweden. The increase of cooling demand occurs mainly during peak demand periods, where there is limited cooling capacity combined with limited distribution capacity in the district cooling network. Sweden has experienced considerable growth in the solar energy market in recent years, though its utilization has been mostly limited to power generation. To fulfill the cooling demand increase, solar driven cooling is a viable solution alternative to traditional cooling methods. The use of solar cooling is still in its early stages in Sweden. The aim of this work is to design a simulation model of a solar absorption cooling system for a full-service restaurant prototype building. The system layout consists of photovoltaic/thermal collector, storage tank, single-effect absorption chiller, auxiliary heater, and cooling tower. The results revealed the system ability to meet the cooling load while delivering sufficient hot water for the establishment. Higher solar fraction confirmed that using photovoltaic/thermal collector is more competitive than solar thermal collectors based on restaurant operational activities. A levelized cost of cooling of 0.164 €/kWh indicated the system cost-effectiveness in comparison to similar setups in other favorable European locations for solar energy utilization.

Keyword: Solar cooling, Absorption cooling, Photovoltaic/thermal collector, TRNSYS

Nomenclature

Symbol	Description	Unit
A	Area	m^2
C	Heat capacity	kJ kg^{-1} K^{-1}
c_p	Specific heat capacity	$Jkg^{-1}K^{-1}$
C_{elec}	Electricity price	€ kWh^{-1}
C_I	Capital investment cost	€
C_{om}	Operating and maintenance cost	€
f	Fraction of rated capacity	−
G	Solar radiation	W m^{-2}
h	Specific enthalpy	kJ kg^{-1}
k	Thermal conductivity	W m^{-1} K^{-1}
L	Distance	m
$\dot{m}$	Mass flow rate	kg s^{-1}
N	Lifespan	year
P	Power	kW
$\dot{Q}$	Heat transfer rate	kW
r	Discount rate	−
T	Temperature	℃
t	Time	s
U	Heat loss coefficient	W m^{-2} K
τ	Transmittance	−
$\dot{W}$	Work	kW
η	Efficiency	−
Subscripts		
aux	Auxiliary	
bot	bottom	
c	Cooling	
chw	Chilled water	
cw	Cooling water	
el	Electrical	
edg	edge	
hw	Hot water	
j	jth node	
out	Outlet	
tank	Storage tank	
th	Thermal	
top	top	
t	Thermal	

Abbreviations		
ANN	Artificial neural network	
COP	Coefficient of performance	
LCOC	Levelized cost of cooling	€ kWh^{-1}
NPC	Net present cost	€
PV/T	Photovoltaic thermal	
SF	Solar fraction	

1. Introduction

The building sector within both the EU and Sweden accounts for about 40% of total energy use (Liu, Rohdin, & Moshfegh, 2015). Space cooling is the fastest-growing use of energy in buildings, both in hot and humid emerging economies where incomes are rising, and in the advanced industrialized economies where consumer expectations of thermal comfort are still growing (IEA, 2018). In Sweden, about 14% of the service sector buildings apply space cooling and about half of these cooling demands were met by district cooling deliveries (Werner, 2017). There are still several potential barriers to installing district cooling in Sweden, such as high upfront costs, a lack of awareness about district cooling systems, and existing infrastructure and property technology limitations. Additionally, legal barriers may also pose challenges for implementing district cooling on a large scale (Palm & Gustafsson, 2018). Cooling demand is expected to continue in the foreseeable future, putting pressure on the country's energy infrastructure. Therefore, finding sustainable and efficient alternative cooling solutions for buildings in Sweden is of importance.

Solar cooling systems offer a sustainable and eco-friendly alternative for air conditioning purposes, as they do not rely on electricity generated from fossil fuels but rather utilize solar energy (Palomba et al., 2017).

A solar absorption cooling system is a type of air conditioning system that uses solar energy to power its cooling cycle. There are several research efforts being conducted on the use of solar absorption chiller systems for cooling applications. Lubis et al. (2016) evaluated the performance of a single-double-effect absorption chiller in tropical Asia regions. During daytime hours, the energy saving could be up to 48% compared with an equivalent vapor compression chiller. Abdullah, Saman, Whaley, and Belusko (2016) investigated the potential of operating a solar-driven absorption chiller for a typical Australian home. The modelling and dynamic simulation of the integrated system

were performed using TRNSYS software and GenOpt ("GenOpt," 2009) was used for optimal sizing of the components. Although the system exhibited a 75% reduction in critical peak power demand, the payback period of the investment was not justified.

Integration of solar absorption systems with other technologies such as energy storage and heat pumps increase the flexibility of the system and making it more practical for use in different applications. Borhani, Kasaeian, Pourmoghadam, and Omid (2023) analyzed the dynamic performance of a photovoltaic/thermal (PV/T) system coupled with an auxiliary heater and an absorption chiller, enabling the generation of electricity, heating, and cooling simultaneously. An artificial neural network (ANN) forecasting model was also developed to predict the system's performance under various climate conditions. Yue et al. (2023) proposed a solar tri-generation supply system that combines a PV/T collector, a heat pump, and an absorption chiller into a single integrated unit. The findings indicated that the system satisfied the load demand of a building in China, achieving an energy efficiency of 32.98% and an exergy efficiency of 17.62%. Moreover, the payback period was estimated to be 7.77 years, which was reasonable in comparison with other conventional systems.

The use of solar absorption cooling in Sweden is currently limited. With the anticipation of increasingly warmer summers in Sweden, the present study aims to assess the feasibility of a solar-assisted chiller powered by PV/T for the purpose of solar cooling. As an added benefit, this system also provides hot water delivery. The motivation behind this dual-purpose system for the restaurant application is threefold. Firstly, it allows the establishment to meet its high temperature hot water needs for sanitization and dishwashing. While district heating is a convenient source of hot water, its supply temperatures cannot meet these specific requirements. Secondly, by utilizing a PV/T system, solar energy is harnessed efficiently for both electricity and heat production, enabling the restaurant to maximize solar energy utilization. Lastly, the system configuration enables heat recovery through a heat exchanger that captures excess heat from the chiller outlet flow to preheat water. Without this heat recovery, the excess heat would otherwise be wasted in the cooling tower discharge.

2. System overview

The schematic of the proposed energy system is shown in Fig. 1. The system comprises a PV/T collector, heat storage tank, auxiliary fluid heater, absorption chiller (AC), and cooling tower, and heat exchanger. The heat storage tank is also equipped with an auxiliary heater in case the solar hot water generation is not sufficient. The AC is a lithium bromide/water single effect-absorption chiller. The PV/T panels produce heat and electricity simultaneously. The thermal storage tank stores hot water from the PVT for hot water usage (state 6) based on the hot water demand profile. When there is a cooling demand, the hot water is transferred to the absorption chiller (state 12). The generated electricity will first be dedicated to powering the chiller and the pumps. The surplus power is then either used for charging the storage tank via an electric coil or in case the tank demand is fully met, it is used in the auxiliary fluid heater to achieve the desired hot water inlet temperature for the chiller (state 12). The energy system in this setup operates without the requirement for battery storage, aiming to maximize cost-effectiveness. The panels and storage unit are appropriately sized to meet a portion of the restaurant's energy demand independently from the heating network and electricity grid. However, the building remains connected to the electricity grid to ensure a continuous supply of energy in case of any shortage. The primary function of the diverter is to control the distribution of hot water flow, effectively directing it to either the storage tank (state 10) or the chiller (state 11). When pump 1 is running, the controller prioritizes supplying water as heat source for the chiller based on the cooling schedule demand. The surplus heat is used for charging the storage tank. The tempering valve splits the input cold water (state 1) between the tee piece 1 and the heat exchanger (state 3). The heat exchanger is used to preheat the cold water before it enters the storage tank (state 4). The source side fluid stream carries the water release from the absorption chiller (state 16) acting as the heat source for the heat exchanger. After cooling down through the heat exchanger (state 17), pump 2 sends the flow back to the tee piece 2 in which it is being mixed with the chiller cooling water (state 15) before delivering it to the cooling tower (state 19). This approach is considered as heat recovery, where the excess heat from the chiller source side is utilized before it is ultimately rejected to the cooling tower. The chilled water is pumped from the absorption chiller unit into a distribution system (state 13). The chilled water carries heat away from the medium and delivers it back to the absorption chiller for the cycle to repeat (state 14).

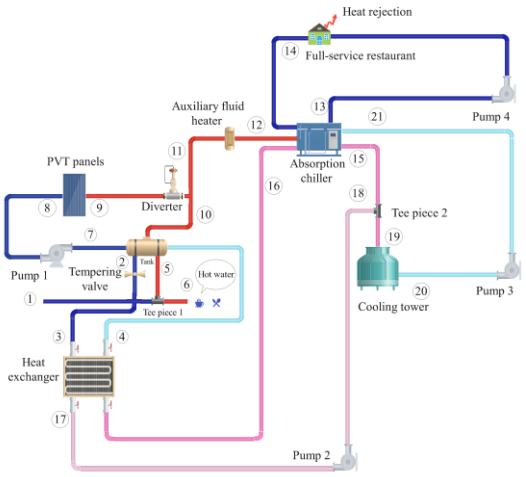


Figure 1. Schematic diagram of the proposed cooling system.

The simulations are performed using TRNSYS software, known for its extensive capabilities in transient modeling of renewable energy systems. All components adopted in this system were provided by the components library. The specification of each component is declared in Tab. 1

Table 1. The components and parameters of the system.

Parameter	Value
PV/T panel	
Collector area	150 m^2
Collector efficiency factor	0.7
Collector plate absorptance	0.9
Collector plate emittance	0.9
Bottom and edge losses coefficient	20 kJ/hm^{2}K
Collector slope	40°
Temperature coefficient of PV cell	-0.0003 1/K
Cell efficiency at reference condition	0.2
Packing factor	0.5
Thermal storage tank	
Tank volume	3 m^3
Tank height	1 m
Top, edge, and bottom loss coefficient	2.5 kJ/hm^{2}K
Absorption chiller	
Rated capacity	80 kW
Rated Coefficient of performance	0.53
Auxiliary electrical power	5 kW

In this work the chilled water is circulated within a full-service restaurant to absorb heat from its desired space. Given the lack of reliable and consistent experimental data from a commercial building with proposed energy system, the authors made the decision to employ a prototype building model as an alternative. The building model used in this study is based on the full-service restaurant model developed by the U.S. Department of Energy (Deru et al., 2011). The building is divided into three thermal zones: kitchen, dining space, and an unconditioned attic. Modifications have been made to the model to represent the building in Sweden (Boverket, 2019). The geometrical and thermal properties of the building are presented in Tab. 2 and Tab. 3, respectively.

Table 2: Geometrical parameters of building.

Parameter	Value	Unit
Area	511	m^2
Floor	Single floor plus attic	-
Floor to ceiling height	3.048	m
South window-wall ratio	28	%
East window-wall ratio	20.22	%
North window-wall ratio	0	%
West window-wall ratio	20.22	%

Table 3. Thermal specifications of model.

Parameter	Value	Unit
Exterior wall	0.29	W/m^2K
Roof	0.12	W/m^2K
Window	2.04	W/m^2K
Interior partition	6.3	W/m^2K

Internal heat loads consist of plug and process loads (157 W/m^2) and lights (20.4 W/m^2). The number of people per floor area is 0.8. Fig. 2 shows the occupancy hourly schedule for weekdays that is used to control the operation of absorption chiller as well.

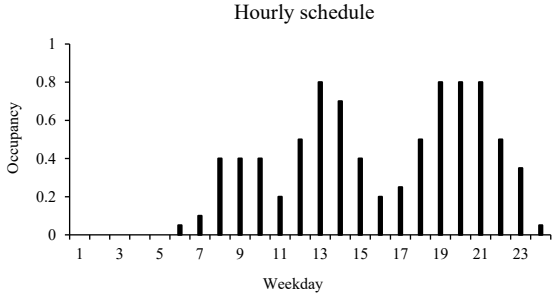


Figure 2. Building occupancy schedules on a weekday.

Fig. 3 gives information about the hourly hot water demand profile of the restaurant based on summer design schedule (Fuentes, Arce, & Salom, 2018; Murakawa, Nishina, Takata, & Tanaka, 2005).

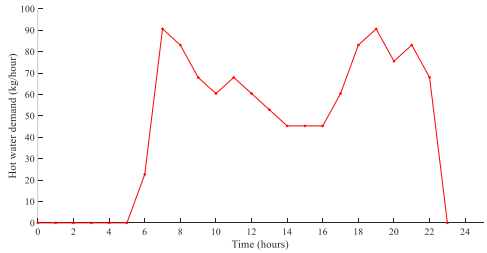


Figure 3. Building hot water draw-off profile.

The primary purpose of the absorption chiller is to meet the cooling requirements of the building. It does so by absorbing heat from the building's cooling load, transferring it to the refrigerant, and ultimately rejecting the heat to the environment through the condenser. The restaurant is divided into two thermal zones, one is the kitchen and the other is the dining area. The cooling setpoint for the dining area is 24°C, while it is 26°C for the kitchen. The cooling load of the restaurant model is shown in Fig. 4. The maximum cooling load is 38.8 kW.

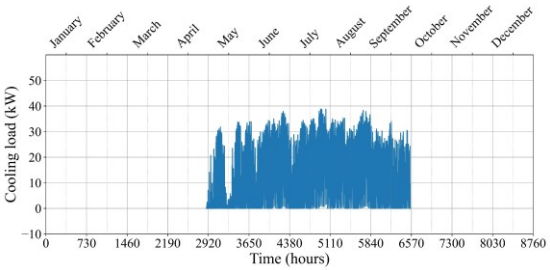


Figure 4. Restaurant hourly cooling load.

3. Modeling

The thermodynamic modeling of the system is divided into multiple subsystems, which include the PV/T panels, heat storage tank, absorption chiller, and the auxiliary components.

3.1. Thermodynamic analysis

The energy assessment aims to analyze the energy conversion quantities, considering the transient behavior of each component. The calculation involves determining the mass and energy balance equations for each component in the system as below (Behzadi & Arabkoohsar, 2020):

$$\sum \dot{m}_{in} = \sum \dot{m}_{out} \quad (1)$$

$$\dot{Q} - \dot{W} = \sum \dot{m}_{out} h_{out} - \sum \dot{m}_{in} h_{in} \quad (2)$$

By applying mass balance and energy balance principles to each component, one can assess its performance.

3.1.1. PVT panels

Th power output from the PV/T panels is directed toward different components based on their electricity demands. The generated electricity is primarily utilized to power the absorption chiller and its associated components such as solution and refrigerant pumps. The surplus power will be used to charge the thermal storage tank via an electrical coil and operating the pumps. If the storage tank is fully charged, the excess power is then utilized by the auxiliary fluid heater.

The electrical and thermal efficiency of the PV/T are as follows:

$$\eta_{el} = \frac{P_{el}}{\tau G A_p} \quad (3)$$

$$\eta_{th} = \frac{Q_u}{\tau G A_p} \quad (4)$$

in which:

$$Q_u = \dot{m}_8 c_p (T_9 - T_8) \quad (5)$$

where η_{el} is the electrical efficiency, G is total incident solar radiation on the collector surface, P is electric power output, τ is the glass cover plate transmittance, $\dot{m}_8$ is water mass flow rate to the panel at state 8, T_9 and T_8 are the outlet and inlet temperature at states 9 and 8, respectively.

The total efficiency of the PV/T panels is determined by adding the electrical efficiency and thermal efficiency as:

$$\eta_{PVT} = \eta_{el} + \eta_{th} \quad (6)$$

3.1.2. Thermal storage tank

The component represents a cylindrical constant volume liquid storage tank with a vertical orientation. It can interact with up to two flow streams and incur thermal losses through the top, bottom, and edges. The tank is divided into temperature nodes to model temperature stratification and each node is governed by the tank energy balance as a function of time (Khan, Badar, Talha, Khan, & Butt, 2018). Through trial-and-error testing of models with varying numbers of nodes, 6 nodes were found to provide the balance between accuracy and computational efficiency. Using more than 6 nodes provided minimal improvements in outlet temperature accuracy while increasing simulation runtimes. The differential equation for the tank nodes can be written as:

$$\frac{dT_{tank,j}}{dt} = \frac{Q_{in,tank,j} - Q_{out,tank,j}}{C_{tank,j}} \quad (7)$$

where $T_{tank,j}$ is temperature of the tank node j, $Q_{in,tank,j}$ and $Q_{out,tank,j}$ is the heat input and output for node j and can be expanded as follows:

$$Q_{in,tank,j} = Q_{aux,t} + \sum \dot{m}_{in} h_{in} \quad (8)$$

$$Q_{out,tank,j} = Q_{loss,top,j} + Q_{loss,bot,j} + Q_{loss,edg,j} + Q_{cond,j} + \sum \dot{m}_{out} h_{out} + Q_{mix,j} \quad (9)$$

where the heat loss from the top, edges, and the bottom of the storage for tank node j are:

$$Q_{loss,top,j} = A_{top,j} U_{top} (T_{tank,j} - T_{env,top}) \quad (10)$$

$$Q_{loss,bot,j} = A_{bot,j} U_{bot} (T_{tank,j} - T_{env,bot}) \quad (11)$$

$$Q_{loss,edg,j} = A_{ed,j} U_{edg} (T_{tank,j} - T_{env,edg}) \quad (12)$$

$A_{top,j}$ represents the tank top surface area for thermal losses (attributed to tank node 1), $A_{bot,j}$ denotes the tank bottom surface area for thermal losses (attributed to tank node N), $A_{edg,j}$ is the tank edge surface area for thermal losses, U_{top} is the storage tank top heat loss coefficient, U_{bot} is the storage tank bottom heat loss coefficient, U_{edg} represents the storage tank edge heat loss coefficient, $T_{tank,j}$ indicates the temperature of individual tank nodes, while, $T_{env,top}$, $T_{env,bot}$, and $T_{env,edg}$ correspond to tank environment temperature for losses through the tank's top, bottom, and edges, respectively.

The nodes in the storage tank can interact thermally via conduction between nodes. The formulation of the conductivity heat transfer from tank node j is:

$$Q_{cond,j} = k_j A_j \frac{T_j - T_{j+1}}{L_j} + k_{j-1} A_{j-1} \frac{T_j - T_{j-1}}{L_{j-1}} \quad (13)$$

where T_j represents the temperature of this node, T_{j+1} is the temperature of the node directly below the current node, T_{j-1} is the temperature of the node directly above the current node; k_j signifies the thermal conductivity of the fluid in node j, k_{j-1} is the thermal conductivity of the fluid in the node directly above the current node; A_j represents the conduction interface area between this node and the one below it, and A_{j-1} is the conduction interface area between this node and the one above it; $L_{cond,j}$ denotes the vertical distance between the centroid of this node and the centroid of the node below, while $L_{cond,j-1}$ represents the vertical distance between the centroid of this node and the centroid of the node above.

$Q_{aux,t}$ is the auxiliary heater input for hot water preparation. At times, the nodes in the storage tank may become thermally unstable (a node has a higher temperature than the node above). If this happens, the model completely mixes any nodes that are unstable at the end of the timestep to avoid problems, $Q_{mix,j}$ is added to the energy balance of the storage tank to account for mixing effects.

3.1.3. Absorption chiller

This component is a single-effect hot-water fired absorption chiller. It is operated using hot water to regenerate the refrigerant in the generator from the refrigerant-absorbent mixture. The component is catalog-based, has its own external file and can predict the chiller's performance within a specified range of input data (Khan et al., 2018). When operating at rated capacity, the design energy input must be provided to the chiller. When the chiller is running at part load, only a fraction of the design energy input is required. With this data, the energy delivered to the chiller by the hot water stream is (Solar Energy Laboratory, 2019):

$$\dot{Q}_{hw} = \frac{\text{Capacity}_{rated}}{\text{COP}_{rated}} f_{DesignEnergyInput} \quad (14)$$

where $\text{Capacity}_{\text{rated}}$ is chiller rated capacity, $\text{COP}_{\text{rated}}$ is the chiller design coefficient of performance, $\text{f}_{\text{DesignEnergyInput}}$ is the fraction of rated capacity required by the cooling machine.

The amount of energy that must be removed from the chilled water stream to bring it from its entering temperature to the setpoint temperature is (Solar Energy Laboratory, 2019):

$$\dot{Q}_{chw} = \dot{m}_{chw} Cp_{chw} \left(T_{chw,in} - T_{chw,set}\right) \quad (15)$$

Where $\dot{\text{m}}_{\text{chw}}$ and Cp_{chw} are the mass flow rate and the specific heat of the chilled water, respectively. The total heat removed from the stream dissipated into the atmosphere in the cooling tower ($\dot{\text{Q}}_{\text{cw}}$) is estimated as (Solar Energy Laboratory, 2019):

$$\dot{Q}_{cw} = \dot{Q}_{hw} + \dot{Q}_{chw} + \dot{Q}_{par} \quad (16)$$

where $\dot{\text{Q}}_{\text{hw}}$ is the energy removed from the hot water stream, $\dot{\text{Q}}_{\text{chw}}$ is the energy removed from the chilled water stream, and $\dot{\text{Q}}_{\text{par}}$ represents the energy consumed by additional components in the system, such as solution pumps, fluid stream pumps, and controls.

It is assumed that the entire energy requirement for auxiliary devices is used whenever the chiller is on, regardless of operating on full or partial load capacity.

3.1.4. Weather data

The hourly weather data of Lund is stored in the TMY-2 (Typical Meteorological Year 2) standard format. Fig. 5 shows the ambient temperature and the global solar radiation in cooling season for Lund, Sweden. The maximum temperature is 28.5 °C, and the horizontal solar radiation reaches above 800 W/m^2 on a summer day.

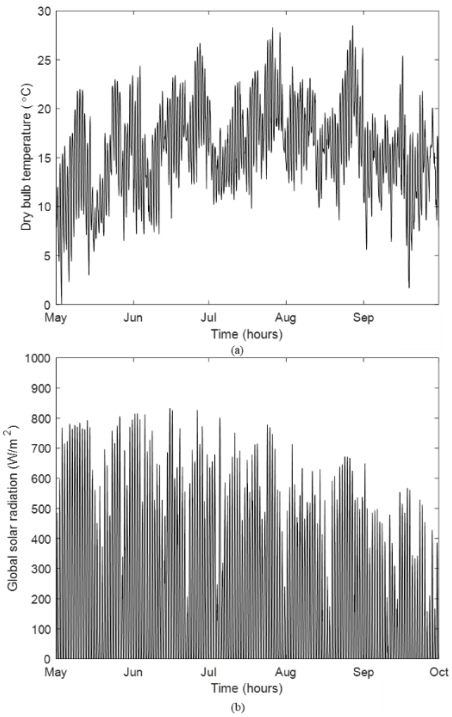


Figure 5. a) ambient temperature and b) global solar radiation on a horizontal plane for Lund in cooling season

3.2. Performance metrics

The evaluation of the performance and effectiveness of an integrated energy system goes beyond solely focusing on efficiency measurements. Therefore, to ensure a holistic approach, additional metrics are analyzed.

Solar fraction is an index used to evaluate the contribution of the solar-driven equipment for cooling and hot water preparation relative to the total energy required to drive the cooling system (Fong, Chow, Lee, Lin, & Chan, 2010).

The solar fraction (SF) is defined as:

$$\text{SF} = \frac{P_{el} + Q_u}{P_{el} + Q_u + Q_{aux,t} + Q_{aux,c}} \quad (17)$$

where P_{el} is the PV/T power output, Q_{u} is the useful energy added by the PV/T to the liquid stream, $\text{Q}_{\text{aux,t}}$ is the auxiliary heater input for hot water preparation, and $\text{Q}_{\text{aux,c}}$ is amount of auxiliary fluid heater input as the energy source for chiller operation.

In addition to the performance metrics, levelized cost of cooling (LCOC) is evaluated as an economic indicator. It represents the price of cooling production and is expressed by (Sajid & Bicer, 2021):

$$\text{LCOC} = \frac{\text{NPC}}{\sum_{i=1}^{25} Q_{cold}} \quad (18)$$

in which Q_{cold} is the amount of heat removal from the building, resulting in a decrease in temperature.

NPC is the net present cost of the system calculated by (Sajid & Bicer, 2021):

$$\mathrm{NPC} = C_I + \sum_{i=1}^{25} \frac{Q_{aux,c} C_{elec} + C_{om}}{(1+r)^i} \quad (19)$$

where r is the discount rate, N is the lifespan of the system, $Q_{aux,c}$ is the auxiliary chiller fluid heater energy consumption, C_{elec} is the unit price of electricity, and C_{om} is the operating and maintenance cost estimated to be 3% of initial investment cost (Sajid & Bicer, 2021).

The investment cost for each of the components is tabulated in Tab. 4. The rest of the key parameters are brought in Tab. 5.

Table 4. Financial specifications of the system

Component	Cost	Reference
Photovoltaic thermal solar collector	$384€/m^2$	(Gu & Zhang, 2021)
Thermal storage tank	$608€/m^3$	(Mortadi & El Fadar, 2022)
Cooling tower	$67.6€/kW$	(Mortadi & El Fadar, 2022)
Single-effect absorption chiller	$300€/kW$	(Saastamoinen & Paiho, 2018)
Inverter	500€	(Reddy Penaka, Kumar Saini, Zhang, & Amo, 2020)
Electric boiler	$78€/kW$	(Zhao, Ge, Sun, Ding, & Yang, 2019)

Table 5. Key input parameters for the simulation.

Parameter	Value	Reference
Electricity price	$0.17€/kWh$	(Gu, Zhang, & Application, 2021)
Discount rate	8%/year	(Gu et al., 2021)
Project lifetime	25 years	(Bellos & Tzivanidis, 2017)

3. Results

The solar system is supposed to be used for cooling as well as hot water preparation. Hence, the scope of simulation is confined from May until the end of September. Using Lund, Sweden as the location of the case study the hot water supply shall be sufficient to satisfy the demands of the establishment. The sanitization and dishwashing, requires higher temperatures for effective cleaning and sterilization so the hot water supply temperature is set to 70°C which is suitable for booster heaters. The tap water temperature is 15 °C. By utilizing the provided information and constructing the model, it becomes possible to generate outcomes that pertain to the performance of the building's energy system throughout the cooling season.

Fig. 6 illustrates the variation of monthly hot water generated during the cooling season. As solar radiation and ambient temperature increase, the rate of energy transfer to the water flowing through the PV/T panels also increases, leading to 313.2 m^3 total produced hot water in June. The monthly produced water cooling is lowest in May and September. For the other months, the amount of water sent toward the chiller is almost constant as the cooling demand rarely has sharp variations in this period.

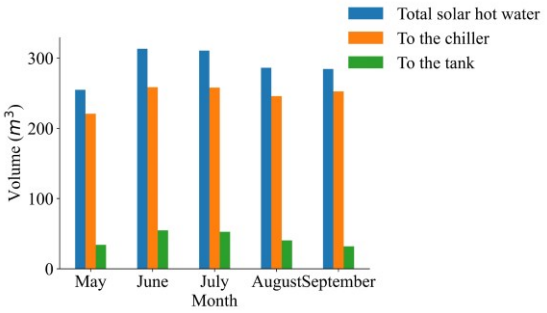


Figure 6. Produced solar hot water volume for each month.

Fig. 7 illustrates the combined heat and electricity produced by the PV/T panels from May to September. It is revealed that the PV/T is more effective at converting solar energy into electricity rather than heat in May and September where there is lower level of solar irradiance. The share of electricity and heat production during the summer months is more balanced which highlights that the PV/T provides a well-rounded energy solution, catering to both electrical and thermal energy needs.

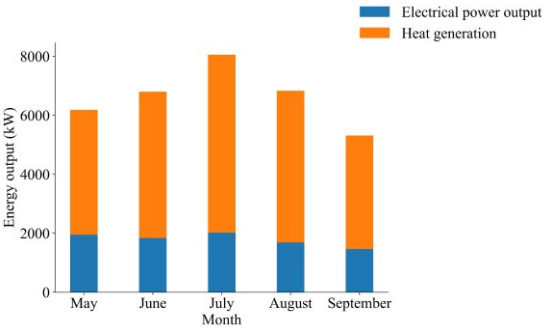


Figure 7. PV/T energy output.

Fig. 8 shows total electricity demand for the operation of the energy system categorized for each component. Among the components, the auxiliary fluid heater has the highest electricity consumption. Due to the lower solar irradiance even in summer months, the thermal energy produced by PVT is generally limited to lower-grade heat, such as hot water preparation. Since the hot water inlet temperature for driving the absorption chiller is 90°C, the additional necessary power required to reach to this setpoint is large relative to other components.

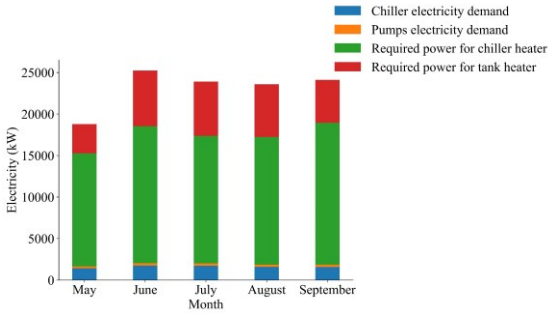


Figure 8. Electricity consumption of each equipment.

A schematic of electricity flows showing possible interactions between the electrical power output from the PV/T and grid electricity to the system is shown in Fig. 9. The total power generated from the PV/T collectors is sufficient to cover the chiller electricity demand. However, the chiller heater relies on grid electricity to meet its power requirement for most of its operation period. It is noticed that only for 16% of the time the reliance on grid is eliminated.

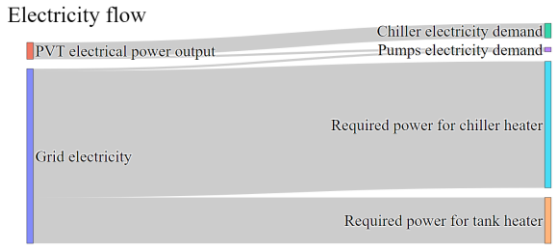


Figure 9. Schematic of electricity flows from PVT and grid to the system.

The cooling generated by the system meets the load requirement, proving the feasibility of the proposed system during the whole simulation period. Besides, it is designed to deliver hot water at a required temperature of 70°C. Fig. 10 shows histograms depicting the distribution of hot water outlet temperatures from the storage tank for each month of the cooling season. The blue bins represent the frequency of hours when the outlet temperature was below the desired setpoint of 70°C. It can be observed that while the average outlet temperature remains close to the setpoint during all months, deviations below the setpoint occur. The percentage of hours below 70°C is higher in the peak summer months of June, July, and August at 49%, 47% and 46% respectively, compared to 23% in May and 34% in September. This occurs as more hot water is diverted to meet the increased cooling demand in summer, causing the tank temperature to drop below the setpoint more often. This is more frequently observed during instances of abrupt changes in water demand, such as in the early morning or after lunch hours. In these scenarios, the recovery time of the auxiliary heater may not be sufficient to keep up with the demand. This can cause the temperature to drop below the setpoint.

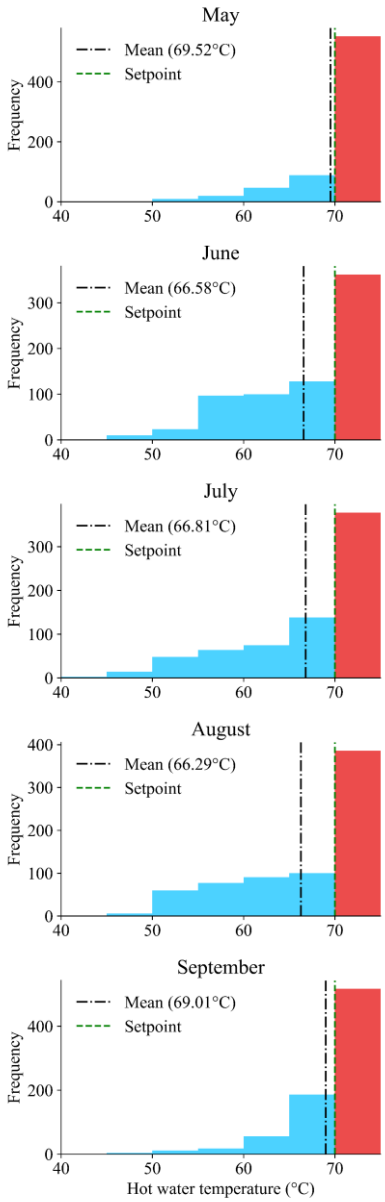


Figure 10. Thermal storage tank outlet temperature violin distribution

Fig. 11 shows the energy performance of the equipment throughout the cooling season. The PVT

has an average thermal efficiency of 28.61% and an electrical efficiency of 6.51%. The single-state absorption chiller average coefficient of performance (COP) is 0.47. From May to August, the solar intensity and ambient temperature increase resulting in higher energy input to the PV/T system and enhanced heat transfer from the PV/T panels and consequently higher thermal efficiency. The system load increases in the summer months. Hence, the absorption chiller may need to operate at higher capacities, which can improve its COP. This is the main reason for the COP drop from August to September.

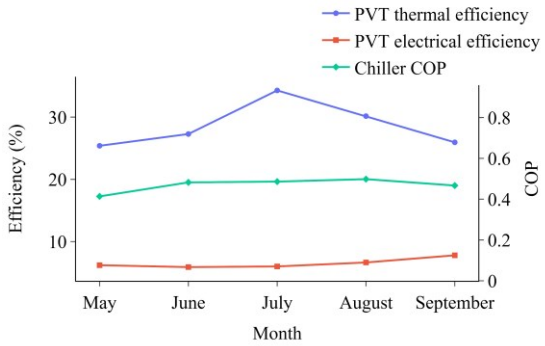


Figure 11. Monthly energy performance of PVT and chiller

Solar-powered integrated energy systems having high solar fractions are more economical (Zhai & Wang, 2009). This metric can be used for comparison between various solar cooling systems. Fig. 12 illustrates the monthly variation of solar fraction throughout the entire cooling season for both the PV/T-based model as well as a model in which the PV/T component is substituted with a solar collector with identical specifications. Another distinction is that the collector-based model is only used for cooling, and hot water supply is discarded. The variation in solar fraction, characterized by higher values in May and September in contrast to the summer months, can be attributed to the reduced cooling demand experienced during these periods for the PV/T-based system. Comparing the annual average values, the solar fraction (SF) for the PV/T-powered system reaches approximately 0.41, which is four times higher than that of the solar collector-based system. Therefore, in solar cooling applications, it is best to integrate them with hot-water supply systems.

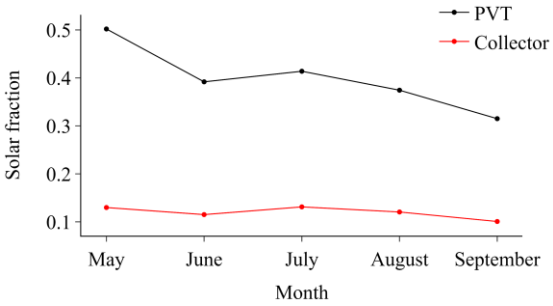


Figure 12. Variation of solar fraction against different months for PV/T and solar collector

To investigate the proposed system from an economic standpoint, the levelized cost of cooling is obtained as 0.164 €/kWh. LCOC provides a standardized approach for comparison among solar thermal cooling systems with absorption chillers. The LCOC values reported by Bellos and Tzivanidis (2017) for solar-thermal absorption chillers for other cities are also depicted in Fig. 13. Generally, locations with higher cooling demand and higher solar radiation attain lower LCOC values. Although the initial investment cost for the proposed PVT-based chiller is higher, the produced cost of cooling is not much higher in comparison with the other cities. It is important to state that the parameter is relatively sensitive to the price of electricity which caused the installation in Madrid with an electricity price of 0.24 €/kWh to be less cost-effective than the present study. Additionally, utilizing PV/T instead of solar thermal collectors offers an added advantage of lowering the levelized cost. This is due to the surplus electricity generated to meet the chiller electricity demand, despite the higher initial investment cost.

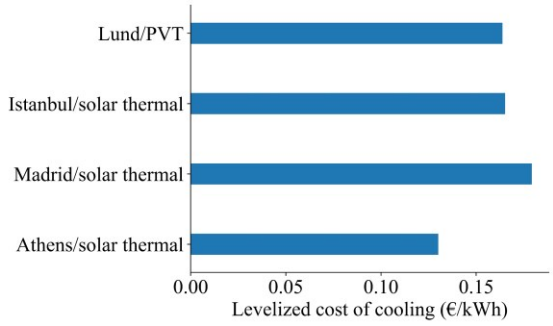


Figure 13. Levelized cost of cooling and installation location for the proposed system and other solar-driven chillers in the literature (Bellos & Tzivanidis, 2017).

4. Summary and Discussions

The cost of cooling is directly linked to the prevailing prices of electricity. Absorption cooling can be increased if there are opportunities to exploit waste heat or renewable heat sources for the process. The feasibility of using a solar-powered absorption chiller for a benchmark model of a full-service

restaurant was assessed. Instead of using the common solar thermal collector, hybrid PV/T collectors were opted for as the hot water production is crucial for a restaurant operation. The system configuration was developed in TRNSYS, and the components were sized based on achieving hot water temperature of 70°C as well as fulfilling the cooling demand. The proposed system met the restaurant's cooling load requirements. It also delivered hot water at the setpoint temperature for many operating hours. However, during peak cooling months the tank outlet temperature deviated below the hot water setpoint more frequently due to abrupt draw increases and limitations in the auxiliary heater recovery time. The surplus electricity generated was found to be sufficient for the chiller operation, provided that all the generated electricity is directed towards it. The PV/T system exhibited an average thermal efficiency of 28.61% and an electrical efficiency of 6.51%. The single-stage absorption chiller achieved an average COP of 0.47. SF and LCOC values confirmed that solar absorption chiller can be a more promising solution in conjunction with PV/T rather than a solar thermal collector in Sweden. This holds particularly true when there is a concurrent demand for electricity and hot water, such as in a restaurant.

Acknowledgment

This research was financially supported by KK-stiftelsen as part of the "SMART- Smart control of district heating networks integrating next generation energy efficient buildings" project.

References

Abdullah, G. F., Saman, W., Whaley, D., & Belusko, M. (2016). Optimization of standalone solar heat fired absorption chiller for typical Australian homes. *Energy Procedia, 91*, 692-701. doi:10.1016/j.egypro.2016.06.232

Behzadi, A., & Arabkoohsar, A. J. E. (2020). Feasibility study of a smart building energy system comprising solar PV/T panels and a heat storage unit. *Energy, 210*, 118528. doi:10.1016/j.energy.2020.118528

Bellos, E., & Tzivanidis, C. (2017). Energetic and financial analysis of solar cooling systems with single effect absorption chiller in various climates. *Applied Thermal Engineering, 126*, 809-821. doi:10.1016/j.applthermaleng.2017.08.005

Borhani, S., Kasaeian, A., Pourmoghadam, P., & Omid, M. (2023). Regional Performance Evaluation of Solar Combined Cooling Heating and Power Systems for Household Demands. *Applied Thermal Engineering, 120666*. doi:10.1016/j.applthermaleng.2023.120666

Boverket. (2019). Boverket´ s Mandatory Provisions and General Recommendations, BBR, BFS 2011: 6 with Amendments up to BFS 2018: 4. In: Boverket Karlskrona, Sweden.

Deru, M., Field, K., Studer, D., Benne, K., Griffith, B., Torcellini, P., . . . Rosenberg, M. (2011). US Department of Energy commercial reference building models of the national building stock.

Fong, K., Chow, T. T., Lee, C. K., Lin, Z., & Chan, L. (2010). Comparative study of different solar cooling systems for buildings in subtropical city. *Solar Energy, 84(2)*, 227-244. doi:10.1016/j.solener.2009.11.002

Fuentes, E., Arce, L., & Salom, J. (2018). 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. doi:10.1016/j.rser.2017.05.229

. GenOpt (Version 3.0.0). (2009). http://simulationresearch.lbl.gov/GO: The Regents of the University of California.

Gu, Y., & Zhang, X. (2021). A Solar Photovoltaic/Thermal (PV/T) Concentrator for Building Application in Sweden Using Monte Carlo Method. In X. Zhang (Ed.), Data-driven Analytics for Sustainable Buildings and Cities: From Theory to Application (pp. 141-161). Singapore: Springer Singapore.

Gu, Y., Zhang, X. J. D.-d. A. f. S. B., & Application, C. F. T. t. (2021). A Solar Photovoltaic/Thermal (PV/T) Concentrator for Building Application in Sweden Using Monte Carlo Method. 141-161.

IEA. (2018). The Future of Cooling. Retrieved from Paris: https://www.iea.org/reports/the-future-of-cooling

Khan, M. S. A., Badar, A. W., Talha, T., Khan, M. W., & Butt, F. S. (2018). Configuration based modeling and performance analysis of single effect solar absorption cooling system in TRNSYS. *Energy conversion management, 157*, 351-363. doi:10.1016/j.enconman.2017.12.024

Liu, L., Rohdin, P., & Moshfegh, B. (2015). Evaluating indoor environment of a retrofitted multi-family building with improved energy performance in Sweden. *Energy and Buildings, 102*, 32-44. doi:10.1016/j.enbuild.2015.05.021

Lubis, A., Jeong, J., Saito, K., Giannetti, N., Yabase, H., & Alhamid, M. I. (2016). Solar-assisted single-double-effect absorption chiller for use in Asian tropical climates. *Renewable Energy, 99*, 825-835. doi:10.1016/j.renene.2016.07.055

Mortadi, M., & El Fadar, A. (2022). Performance, economic and environmental assessment of solar cooling systems under various climates. *Energy Conversion and Management, 252*, 114993. doi:10.1016/j.enconman.2021.114993

Murakawa, S., Nishina, D., Takata, H., & Tanaka, A. (2005). An analysis on the loads of hot water consumption in the restaurants. Paper presented at the Proceedings of the 31st W062 International Symposium on Water Supply and Drainage for Buildings, Brussels, Belgium.

Palm, J., & Gustafsson, S. (2018). Barriers to and enablers of district cooling expansion in Sweden. *Journal of Cleaner Production, 172*, 39-45. doi:10.1016/j.jclepro.2017.10.141

Palomba, V., Vasta, S., Freni, A., Pan, Q., Wang, R., & Zhai, X. (2017). Increasing the share of renewables through adsorption solar cooling: A validated case study. *Renewable Energy, 110*, 126-140. doi:10.1016/j.renene.2016.12.016

Reddy Penaka, S., Kumar Saini, P., Zhang, X., & Amo, A. d. (2020). Digital mapping of techno-economic performance of a water-based solar photovoltaic/thermal (PVT) system for buildings over large geographical cities. *Buildings, 10(9),* 148. doi:10.3390/buildings10090148

Saastamoinen, H., & Paiho, S. (2018). Prospects for absorption chillers in Finnish energy systems. *Energy Procedia, 149*, 307-316. doi:10.1016/j.egypro.2018.08.194

Sajid, M. U., & Bicer, Y. (2021). Comparative life cycle cost analysis of various solar energy-based integrated systems for self-sufficient greenhouses. *Sustainable Production and Consumption, 27,* 141-156. doi:10.1016/j.spc.2020.10.025

Solar Energy Laboratory, T. T. E. G., CSTB (Centre Scientifique et Technique du Bâtiment). (2019). In TRNSYS 18 a TRaNsient SYstem Simulation program, Mathematical Reference (Vol. 4, pp. 149-156). United States.

Werner, S. (2017). District heating and cooling in Sweden. *Energy, 126*, 419-429. doi:10.1016/j.energy.2017.03.052

Yue, H., Xu, Z., Chu, S., Cheng, C., Zhang, H., Chen, H., & Ai, D. (2023). Study on the Performance of Photovoltaic/Thermal Collector–Heat Pump–Absorption Chiller Tri-Generation Supply System. *Energies, 16(7)*, 3034. doi:10.3390/en16073034

Zhai, X., & Wang, R. (2009). A review for absorbtion and adsorbtion solar cooling systems in China. *Renewable and Sustainable Energy Reviews, 13(6-7)*, 1523-1531. doi:10.1016/j.rser.2008.09.022

Zhao, S., Ge, Z., Sun, J., Ding, Y., & Yang, Y. (2019). Comparative study of flexibility enhancement technologies for the coal-fired combined heat and power plant. *Energy Conversion and Management, 184,* 15-23. doi:10.1016/j.enconman.2019.01.030

In-Depth System-Level Energy Analysis of Hybrid Electrified Commuter Aircraft for Improved Energy Efficiency

Dimitra-Eirini Diamantidou[a,*] Valentina Zaccaria[a] Anestis Kalfas[b]

[a]*Department of Sustainable Energy Systems, Mälardalen University, Västerås SE-72220, Sweden* [b]*Department of Mechanical Engineering, Aristotle University of Thessaloniki, Thessaloniki GR-54124, Greece*
[*]dimitra.eirini.diamantidou@mdu.se

Abstract

This work presents a comprehensive analysis of hybrid electric propulsion systems in commuter aircraft, aimed at enhancing energy efficiency. The study utilizes an aircraft conceptual design library, OpenConcept, to perform evaluations of various aircraft components and their interrelationships. The methodology integrates aerodynamics, propulsion, and mission analysis within a common framework to optimize the aircraft design. The analysis focuses on a 19-passenger commuter aircraft, employing a series/parallel hybrid-electric architecture. The gradient-based Sequential Least Squares Programming optimizer is utilized to optimize design variables such as battery weight, engine power, and the selected power ratios, while adhering to operational constraints. Through a rigorous Design of Experiments study, the paper highlights that even when considering the current battery technology, hybrid-electric propulsion yields substantial energy savings for short-haul missions. The fuel and energy consumption reductions are evident, particularly for shorter ranges. However, for extended missions, the critical role of advanced battery energy density is emphasized to achieve significant energy efficiency improvements.

1 Introduction

Aerospace sustainability has received considerable focus over the past years. Aviation's environmental impact, particularly greenhouse gas emissions and noise pollution are the main drivers of making aviation more efficient and, hence, more sustainable. This is reflected by the sustainability goals (Darecki et al., 2011; Mangelsdorf, 2012) set by various organizations across the globe. Despite the operation disruptions due to the pandemic, the sector is gradually returning to normal operations and is expected to exceed pre-pandemic levels (IATA, 2022). Therefore, there is a need for more sustainable aircraft in the coming future.

Electrified propulsion is a promising technology which is receiving increasing attention in the last decades. Due to the current energy storage technology and inherit additional weight, fully-electric propulsion is only feasible for limited range. Hybrid-electric propulsion systems are considered as the stepping-stone towards zero-emission aircraft. These aircraft concepts combine traditional gas turbines with electrical motors and energy storage systems. They present potential to reduce fuel consumption, emissions and improve the overall system energy efficiency (Felder, 2016; Pornet & Isikveren, 2015).

The electrification of aircraft industry focuses on different market segments. One of them is the short-haul segment, where light commuter aircraft applications are covering distances of maximum 600-800 nautical miles. Such applications use turboprops and small regional jets with turboprops prioritizing fuel efficiency while turbofans and turbojets offer higher cruise speeds.

Several benefit estimations have been presented in the open literature based on conceptual design of hybrid-electric aircraft. As it was reported by (Kruger et al., 2018), different architectures have different optimal applications with hybrid-electric being suitable for intermediate ranges ($\leq$ 800 NM). This work was extended with the mission analysis of commuter aircraft (Kruger & Uranga, 2020) and it was presented that hybrid-electric configuration can lead to a 63% reduction in energy consumption, with a cost of 52% increase in the take-off weight. (Zamboni et al., 2019) analysed different hybrid-electric architectures with the series/parallel one emerging as the most promising due to the combination of benefits in both aerodynamic performance and propulsion system efficiency. The authors reported a 28% and 14% reduction in fuel and energy consumption respectively compared to the baseline aircraft considered. In addition, the ELICA EU-funded project reported a 56% reduction in total energy consumed for a series/parallel partial hybrid commuter aircraft with Entry-Into-Service (EIS) 2025 when compared to a reference aircraft with EIS 2014 (Nicolosi et al., 2022). Finally, (Schäfer et al., 2019)

estimated that short-haul electrified aircraft has the potential to replace up to 15% of global revenue passenger kilometers and a substantial number of global departures.

The addition of extra electrical components can lead to increased complexity to the system, requiring a thorough evaluation of its effects. The generation of thrust relies on two sources: traditional fuel and electrical energy from hybrid systems. Deciding when to use each energy source involves various factors and needs in-depth investigation. Many design choices depend the components' efficiency, weights, and interrelations (Moore, 2014). Instead of the traditional approach of designing aircraft, there's a growing need to shift towards more comprehensive methodologies. These combine exploring different design possibilities with optimization techniques while taking into account a range of aspects from different areas of aircraft design. This shift is based on the understanding that the connections between these different disciplines have considerable effects on the system performance (Martins & Lambe, 2013). Therefore, there is a requirement for flexible and efficient design tools that can handle various aspects of aircraft design and their complex relationships. Such tools allow for a holistic approach that not only considers the efficiency of individual parts but also takes into consideration the broader effects and interactions within the entire system.

This study delves into the advantages presented by a series/parallel partial hybrid-electric concept designed for a 19-passenger commuter aircraft. The architecture incorporates a conventional turboprop engine and a motor-driven e-propeller per wing. In addition, an on-board battery system is integrated while the turboprop engines coupled with generators. This arrangement enables the e-propeller to operate using either battery or engine-generated energy. To facilitate these investigations, a flexible aircraft conceptual design tool, built upon the OpenConcept library (Brelje & Martins, 2018), is developed. Notably, essential adaptations to the computational scheme are introduced, enabling fast and approximate calculations crucial for mission performance analysis. By integrating different disciplines, the study paves the way for the design and operation of hybrid-electric aircraft, with a strong focus on advancing fuel and energy efficiency. By employing a comprehensive design of experiments (DOE) approach, the research focuses on two pivotal variables: the mission range (measured in nautical miles) and the battery energy density (Wh/kg). These key factors are strategically selected for their profound influence on the aircraft's operational capabilities and overall efficiency. A comprehensive understanding of the interplay between these variables facilitates an assessment of the feasibility to achieve the desired mission range while optimizing the battery offering the potential to elevate the aircraft's holistic performance.

2 Methodology

2.1 Aircraft conceptual design framework

The present work employs a general purpose aircraft design toolkit which includes different aircraft component models to perform individual calculations while taking into consideration their interrelations. The framework is based on the OpenConcept developed by (Brelje & Martins, 2018) which is an open-source python library. OpenConcept is an adaptable, low-fidelity aircraft design library and its main purpose is to provide fast mission results for aircraft conceptual design.

The user provides a set of parameters to the library; aircraft geometry characteristics, typical aircraft weights such as the maximum take-off weight (MTOW), propulsion component characteristics and finally mission profile parameters for all mission phases. The library uses the set of input parameters and the pre-defined mission to perform weight estimations, basic aerodynamic calculations and finally mission performance analysis to compute high-level variables such as fuel and energy consumed. It employs a non-linear Newton solver in order to set an appropriate lift coefficient and throttle of different powertrain components to satisfy the pre-defined mission requirements.

The OpenConcept library has a variety of different size aircraft models available as well as series hybrid-electric and parallel hybrid-electric modelling capabilities (Adler et al., 2022; Fouda et al., 2022). It is important to note that the weight estimation calculations are based on textbook calculations using empirical formulas (Raymer, 2018; Roskam, 2019; Torenbeek, 2013) and different calculations are implemented based on the aircraft class considered. Furthermore, the employed library is built on top of the OpenMDAO framework and is extensively discussed in the work of (Gray et al., 2019). Any of the aforementioned input variables can be used as an optimization design variable, and, hence, the analysis and optimization of the aircraft system is enabled.

Within the present work, a conventional 19-passenger aircraft model has been developed to enable the investigation of hybrid-electric commuter aircraft concepts. The conventional model is based on the Beechcraft 1900D aircraft featuring two turboprop engines. The main aircraft design parameters, such as the wing area and Operating empty weight (OEW) are matched with the publicly available data for the Beechcraft 1900D (Beech Commuter Airliners, 2000). The aircraft model matches the value of OEW with a very small deviation of 0.04%. This ensures that the model

correlations can closely capture the reference aircraft. Finally, it is important to note that the conventional Beechcraft 1900D aircraft serves as the reference aircraft throughout this work.

2.2 Series/parallel partial hybrid-electric architecture

The work focuses on the series/parallel partial hybrid-electric architecture. The Therefore, a hybrid-electric aircraft model has been developed for the purposes of this work using the OpenConcept library. The model is based on the conventional 19-passenger Beechcraft 1900D model mentioned in the previous subsection.
A schematic of the propulsion model used is presented in Figure 1. The powertrain architecture includes two turboprop conventional engines, a battery system, a Power Management and Distribution (PMAD) system and two e-propellers driven by electrical motors. The PMAD system shown in Figure 1 is introduced in the model to ensure a consistent power supply from different energy sources. The powertrain architecture has been modified accordingly in the model to include the required components.

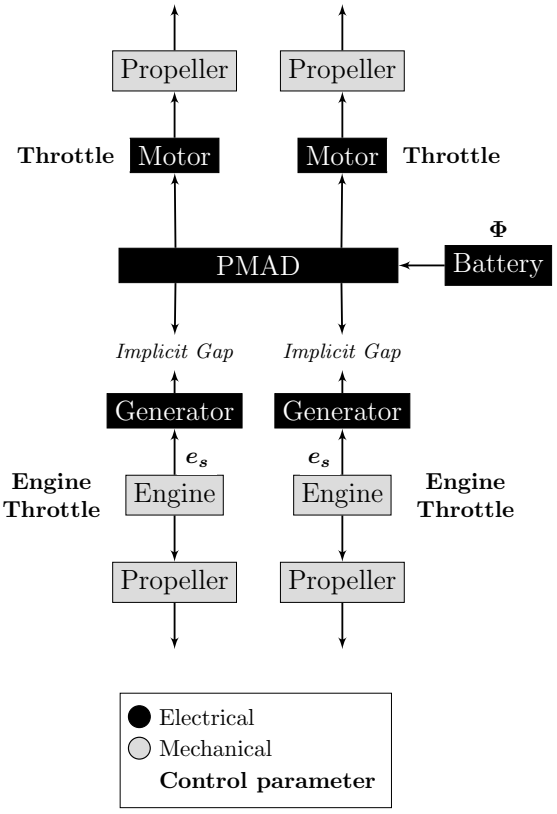


Figure 1. Schematic of a twin-engine series/parallel partial hybrid electric propulsion model in OpenConcept.

As shown in Figure 1, the computational scheme is updated as well, as both thermal and electrical components can contribute to the thrust generation in this concept. The turboprop engines can deliver power to both the propeller (denoted by P_{prop}) and the generator (denoted by P_{hybrid}). Therefore an engine power split is defined within the model as shown here:

$$e_S = \frac{P_{prop}}{P_{prop} + P_{hybrid}} \tag{1}$$

The e_S is set by the user (or optimizer) and is kept at the same level for the two engines to achieve thrust balance. Furthermore, it can take different values for different mission phases but is selected to remain constant throughout each phase.
In addition, to determine the battery contribution on the thrust produced by the e-propellers, the supply power ratio Φ is introduced in the PMAD system, following the definition presented by (Isikveren et al., 2014; de Vries et al., 2019):

$$\Phi = \frac{P_{bat}}{P_{bat} + P_{gen,1} + P_{gen,2}} \tag{2}$$

where, P_{bat} is the power delivered by the battery and P_{gen} is the power delivered by the generator to the PMAD. To be able to match the power requirement from the motors, an implicit gap is exposed to the Newton solver. The engines' throttle is set by the solver in order to achieve the PMAD power requirement P_{hybrid}, whereas the Φ is set by the user. Finally, it is important to highlight that the thrust split between the mechanical and electrical propeller is not set directly but is a result of the power calculated by the combination of e_S and Φ. The motors' throttle is set again by the solver in order to achieve steady flight conditions as it will be discussed in subsection 2.3.
The main model assumptions for individual powertrain components are summarized in Table 1. The time-frame under consideration for the entry into service (EIS) of the concept is set to 2035. This is reflected on the assumptions for the electrical component efficiencies. For the turboprop engines, the model performance characteristics are based on the Pratt & Whitney PT6A-67D unit (Badger et al., 1994).

2.3 Mission analysis

The nominal mission profile selected is presented in Figure 2. The mission analysis includes a main mission, where the high-level parameters are extracted and a diversion mission of 100 nautical miles to consider the scenario in which the aircraft must land in a different airport. Finally, according to the CS-23 certification requirements (EASA, 2018) for 19 passenger commuter aircraft, the aircraft shall be able to fly for 30mins at 1500ft using only one energy source. This requirement is included in the mission analysis and aircraft design as a loiter phase.

Table 1. Powertrain technology assumptions (EIS 2035).

Parameter	Value
P&W PT6A-67D	2×950 kW
PSFC*	0.074/0.094/0.1* g/(kW*s) (climb/cruise/loiter)
Motor power density	10 kW/kg
Generator power density	10 kW/kg
Motor efficiency	0.95
Generator efficiency	0.96
Battery efficiency	0.97
PMAD efficiency	0.99

* *estimation*

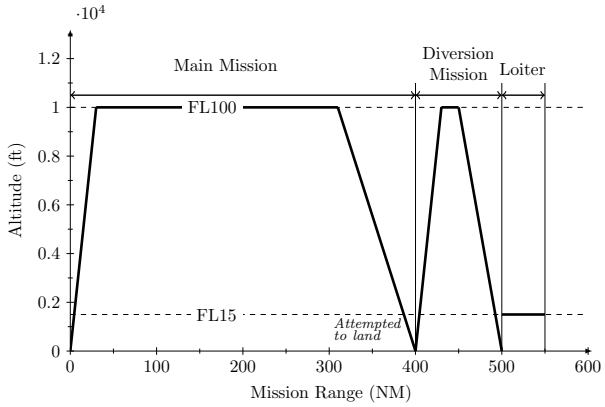


Figure 2. Typical mission profile for commuter aircraft.

Each mission phase is split into segments by using intermediate points, in which the vertical and horizontal speed are pre-defined. For each flight segment, the library calculates the value of the residual R_{thrust} using the following equation:

$$R_{thrust} = T - D - mg\sin\gamma \tag{3}$$

where, T is the thrust, D is the drag, m is the current aircraft weight, g is the gravitational acceleration and finally γ is the aircraft angle. The OpenMDAO's Newton solver is responsible to estimate the primary thrust control parameter (engine or motor throttle) in order to achieve zero horizontal acceleration. Furthermore, in the climb and descent phases the excess thrust or drag respectively are achieved in such a way to match the vertical speed, whereas in the case of cruise and loiter, the thrust and drag forces should be equal. Finally, for each segment, the aircraft weight is updated based on the fuel consumed in the previous segment. Parameters such as fuel and energy consumed are integrated with respect to time using the Simpson's rule. The diversion mission and loiter phase fuel and energy are not directly presented in this work, however they do affect the aircraft design as according to the CS-23 certification requirements, the maximum MTOW is 8618 kg. The model calculations are considering that enough fuel or electrical energy is on board to satisfy the main mission along with diversion mission and loiter.

Finally, the top level aircraft requirements (TLARs) selected for the concept under investigation are summarized in Table 2. The TLARs of the concept are based on the limitations imposed by the CS-23 certification. The payload is calculated based on the numbers of passengers, passenger weight and baggage weight as shown in 2.

Table 2. TLARs for hybrid-electric concept.

Parameter	Value
MTOW	$\leq$ 8618 kg
Number of passengers	19
Payload	1881kg
	19*(87 + 12)
Nominal Mission	740.8 km (400 NM)
Cruise altitude	3048 m (10000 ft)
Loiter altitude	457.2 m (1500 ft)
Cruise Mach number	0.35
Rate of climb (MTOW, SL, ISA)	$\geq$ 6.35 m/s
Approach speed	$\leq$ 62 m/s

2.4 Optimization problem

The OpenConcept library is developed to enable for conceptual design optimizations. In this work, the hybrid-electric concept is optimized for different mission ranges and battery technology assumptions. The Sequential Least Squares Programming approach (Bonnans et al., 2006) is used to size the propulsion system components for minimum fuel burn on the selected mission.

The optimization problem is presented in Table 3. The MTOW, battery weight, supply power ratio Φ and engine power split e_S are varied in order to minimize the block fuel. The block fuel in this study denotes the main mission fuel. For Φ, only the climb, cruise and descent phases are considered and therefore number of design variables for Φ is 3. The diversion mission and loiter Φ are not considered directly in the optimization and assumed to be zero. For the diversion mission, this choice is made to ensure that the fuel consumption and therefore environmental benefit takes place during the main mission, which is the primary flight scenario in most cases. However, the diversion mission fuel consumption has an effect on the optimization results mainly due to the fact that MTOW contains the total fuel and the MTOW is limited to 8618kg. For the loiter phase, this choice is aligned with the certification requirements.

On the other hand, the e_S is selected to be varied by the optimizer for all mission phases included in the study. Therefore, the final count of design variables from e_S in the optimization problem is 7. This choice is made to ensure safe operation for all powertrain components during the whole mission. Finally, the rated power for the generator and motor are varied to size the components based on the maximum power they can deliver during the mission.

Table 3. Optimization problem definition for hybrid-electric aircraft.

	Variable	Quantity
minimize:	Block fuel	1
by varying:		
	MTOW	1
	W_{batt}	1
	P^*_{motor} (rated)	1
	P^*_{gen} (rated)	1
	Φ	3
	e_s	7
subject to:		
	$0 \leq R_{MTOW} \leq 5$	1
	$0.2 \leq SOC_{batt,loiter}$	1
	$0.1 \leq throttle \leq 1$	7
	$0.1 \leq eng\ throttle \leq 1$	7

After the weight calculations and mission analysis take place within the library, the residual MTOW is calculated as shown here:

$$R_{MTOW} = (MTOW - OEW - W_{fuel} - W_{batt} - W_{payload}) \quad (4)$$

where, $MTOW$ is a design variable, OEW is calculated, W_{fuel} is integrated from the fuel flow rate with respect to time for the whole mission, W_{bat} is a design variable and $W_{payload}$ is set by the TLARs in Table 2. The optimizer constrains the R_{MTOW} to small values (between 0 and 5kg). Furthermore, the state of charge (SOC) of the battery at the loiter phase (end of mission) shall be greater than 20% in order to make sure that the battery is not fully discharged at the end of the mission and ensure its safe operation. Finally, the motors' and engines' throttles are constrained to ensure their realistic operation within the flight envelope. The component throttles are defined as the ratio between the power delivered by the component and the maximum (rated) power of the component.

2.5 Design of Experiments (DOE)

A DOE study is conducted considering two variables, the main mission range and the battery energy density. The selection of the main mission range variable aims to explore various aircraft operations and assess the potential advantages of introducing commuter aircraft for short-haul flights. The battery specific energy density variable is closely tied to the assumptions regarding technology. By examining different levels of energy density, valuable conclusions can be drawn regarding the potential of electrified aircraft technology, particularly for aircraft falling under CS-23 certification.

A full factorial design with 20 levels, and, hence 20^2 samples, is selected for this study. For each combination of mission range and battery energy density, an optimization is conducted as presented in Table 3 in order to minimize the block fuel. A uniform distribution of the examined points in the design space is selected, with optimizations of different range and energy density taking place for each sampling point.

3 Results

The results for the MTOW and battery weight ratio are depicted in Figures 3 and 4 respectively. The results confirm that electrifying short-haul aircraft leads to a significant challenge in terms of MTOW. The upper limit of 8618 kg on MTOW imposed by the CS-23 certification, is reached for the majority of optimized designs. Lowering the MTOW values would require disruptive advancements in battery technology, especially for longer missions.

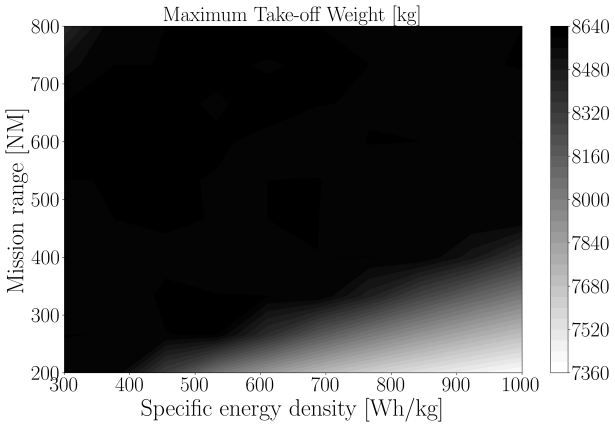


Figure 3. Maximum Take-off weight (MTOW).

Figure 4 shows that for shorter missions as the battery specific energy density increases, the available battery on board is reduced. This is expected as for higher energy density, the required energy can be achieved for lower weight. On the contrary, for the longer missions, a different trend is observed. As the battery energy density decreases, the available battery weight starts decreasing. This is a result of the MTOW upper limit. For longer missions, the upper limit of MTOW weight is reached and therefore, it affects the amount of available battery weight on board.

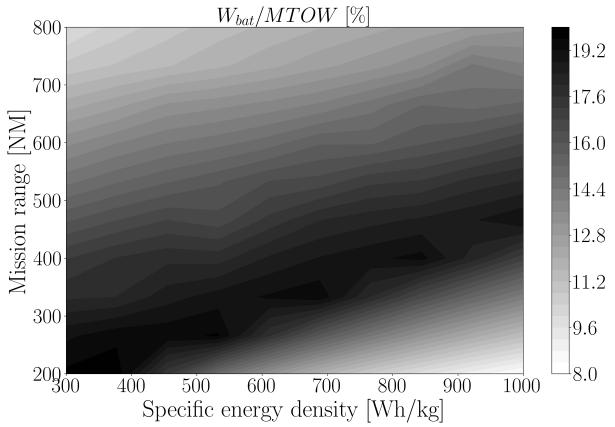


Figure 4. Battery weight on board divided by the MTOW.

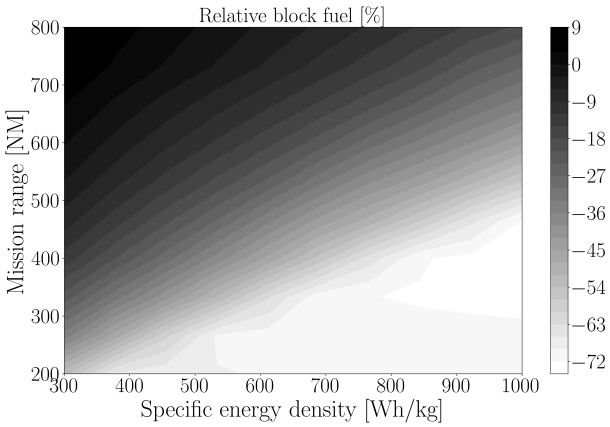


Figure 5. Relative block fuel burn compared to the conventional aircraft.

The energy consumption aspects of the optimized design are also investigated. The block fuel and total energy results are normalized against the corresponding values per mission range derived from the conventional aircraft model as discussed in subsection 2.1. The equation below is used for both quantities of interest:

$$Q_{hybrid}[\%] = \frac{Q_{hybrid} - Q_{conv}}{Q_{conv}} * 100\% \qquad (5)$$

It is important to note that for the hybrid-electric aircraft, both electrical and fuel energy are considered in the total energy consumption whereas for the conventional aircraft the lower heating value of the Jet-A fuel is used in order to calculate the energy from the fuel consumed. The lower heating value of the fuel is assumed to be 42.8 MJ/kg. The relative results are presented in Figures 5 and 6. It is evident from the graphs, that the series/parallel hybrid-electric aircraft can achieve not only lower fuel consumption but also lower total energy consumption. This improvement is more substantial for higher battery energy density, as expected. However, for short mission ranges a considerable improvement is shown, even with the current battery technology (300 Wh/kg).
Finally, the ratio between the electrical and total energy is shown in Figure 7. This ratio is the degree of hybridization (DoH) for hybrid-electric aircraft concepts. It is important to note that when increased electrical energy consumption is enabled, the total energy consumption is reduced. Therefore, Figures 6 and 7 have opposite trends. This is because of higher efficiencies achieved through the introduction of electrical components. This observation confirms the benefits of high DoH in short-haul aircraft applications.

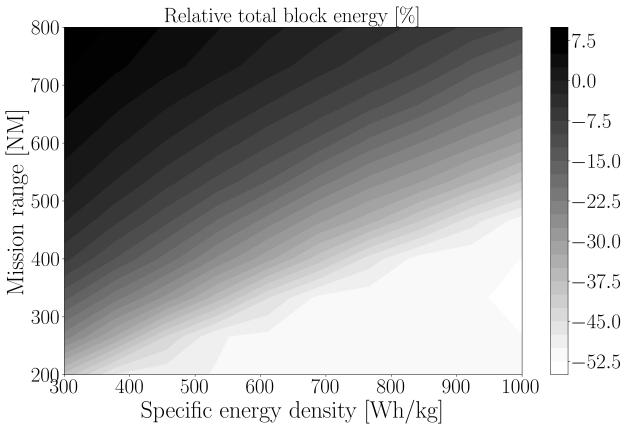


Figure 6. Relative block total energy compared to the conventional aircraft.

4 Summary and Discussions

The investigation of the advantages offered by the series/parallel partial hybrid-electric concept for a 19-passenger commuter aircraft is the focus of this work. To achieve this objective, a novel aircraft conceptual design framework is developed and thoroughly presented, leveraging the OpenConcept library. The presented framework integrates a variety of aircraft disciplines within the design process. The hybrid-electric aircraft model, based on the well-established Beechcraft 1900D aircraft, undergoes tailored computational developments to consider pivotal control parameters such as the supply power ratio (Φ) and engine power split (e_S). This modification enables the analysis and optimization of the aircraft's application. The central focus of this study lies in mission analysis and the system performance evaluation, considering a typical commuter aircraft mission profile. A comprehensive examination of both main and diversion missions, along with a dedicated 30-minute loiter phase, sheds light on critical high-level system variables, including fuel and energy consumption. The overall optimization objective revolves around optimizing the size of propulsion system components to minimize fuel consumption during the main mission.

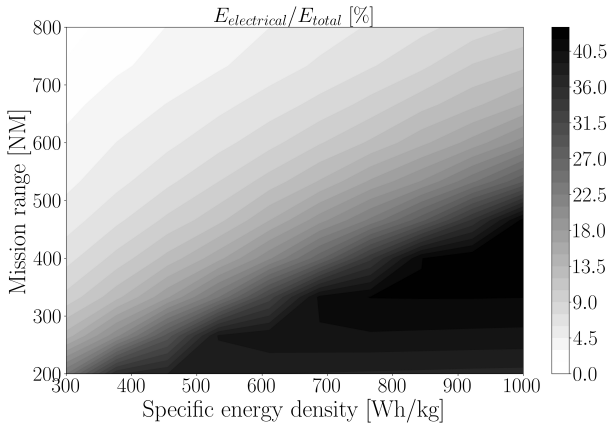


Figure 7. Electrical block energy consumed.

The findings of this study validate the benefits of the series/parallel partial hybrid-electric architecture. The optimized designs present considerable reductions in fuel burn and energy consumption. Notably, the study highlights the importance of improved battery technology in achieving substantial reductions in fuel and energy consumption for long-range missions, while also pointing out the advantages of a range of battery energy densities for shorter missions.

This work fundamentally highlights the significance of considering interdisciplinary design tools and optimization methodologies. As the aviation industry strives for enhanced sustainability, the integration of hybrid-electric propulsion systems into short-haul aircraft emerges as a straightforward way to achieve reductions in both energy and fuel usage.

Future work should include the integration of a detailed engine model for a more realistic estimation of fuel consumption when the engine is either redesigned or working in part-load conditions. In addition, a detailed investigation of the optimized power management for a range of different battery technology projections and mission ranges will provide valuable insights into the trade-offs in hybrid-electric aircraft systems for short-haul operation.

References

Adler, E. J., Brelje, B. J., & Martins, J. R. R. A. (2022). Thermal management system optimization for a parallel hybrid aircraft considering mission fuel burn. *Aerospace*, *9*(5). doi: 10.3390/aerospace9050243

Badger, M., Julien, A., LeBlanc, A. D., Moustapha, S. H., Prabhu, A., & Smailys, A. A. (1994). The PT6 Engine: 30 Years of Gas Turbine Technology Evolution. *Journal of Engineering for Gas Turbines and Power*, *116*(2), 322-330. doi: 10.1115/1.2906823

Beech Commuter Airliners. (2000). *Beech 1900 airliner pilot training manual*. Flushing, New York: FlightSafety International, Inc.

Bonnans, J.-F., Gilbert, J. C., Lemaréchal, C., & Sagastizábal, C. A. (2006). *Numerical optimization: Theoretical and practical aspects*. Springer Berlin, Heidelberg. doi: 10.1007/978-3-540-35447-5

Brelje, B. J., & Martins, J. R. R. A. (2018). Development of a conceptual design model for aircraft electric propulsion with efficient gradients. In *Proceedings of the aiaa/ieee electric aircraft technologies symposium.* Cincinnati, OH: AIAA. doi: 10.2514/6.2018-4979

Darecki, M., King, I., Edelstenne, C., Ky, P., Enders, T., Mathieu, M., ... Wörner, J. D. (2011). *Flightpath 2050:Europe's Vision for Aviation: Report of the High-Level Group*. Publications Office of the European Union, Luxembourg.

de Vries, R., Brown, M., & Vos, R. (2019). Preliminary sizing method for hybrid-electric distributed-propulsion aircraft. *Journal of Aircraft*, *56*(6), 2172-2188. doi: 10.2514/1.C035388

EASA. (2018). *Easy access rules for normal, utility aerobatic and commuter category aeroplanes (cs-23) (initial issue).* Brussels, Belgium: European Aviation Safety Agency. (Acccesed on May 29, 2023)

Felder, J. L. (2016). *NASA electric propulsion system studies* (Tech. Rep. No. GRC-E-DAA-TN28410). National Aeronautics and Space Administration (NASA).

Fouda, M. E. A., Adler, E. J., Bussemaker, J., Martins, J. R., Kurtulus, D., Boggero, L., & Nagel, B. (2022). Automated hybrid propulsion model construction for conceptual aircraft design and optimization. In *33rd Congress of the International Council of the Aeronautical Sciences, ICAS 2022.*

Gray, J. S., Hwang, J. T., Martins, J. R. R. A., Moore, K. T., & Naylor, B. A. (2019). OpenMDAO: An Open-Source Framework for Multidisciplinary Design, Analysis, and Optimization. *Structural and Multidisciplinary Optimization*, *59*, 1075-1104. doi: 10.1007/s00158-019-02211-z

IATA. (2022). *Global outlook for air transport: Sustained recovery amidst strong headwinds.* `https://shorturl.at/qwBL6`. (Last accessed on May 29, 2023)

Isikveren, A., Kaiser, S., Pornet, C., & Vratny, P. (2014). Pre-design strategies and sizing techniques for dual-energy aircraft. *Aircraft Engineer-*

ing and Aerospace Technology, *86*(6), 525-542. doi: 10.1108/AEAT-08-2014-0122

Kruger, M., Byahut, S., Uranga, A., Gonzalez, J., Hall, D. K., & Dowdle, A. (2018). Electrified aircraft trade-space exploration. In *2018 aviation technology, integration, and operations conference.* doi: 10.2514/6.2018-4227

Kruger, M., & Uranga, A. (2020). The feasibility of electric propulsion for commuter aircraft. In *Aiaa scitech 2020 forum.* doi: 10.2514/6.2020-1499

Mangelsdorf, M. (2012). *Environmentally Responsible Aviation N+2 Advanced Vehicle Concepts NRA Status* (Tech. Rep. No. DFRC-E-DAA-TN3702). National Aeronautics and Space Administration (NASA).

Martins, J. R. R. A., & Lambe, A. B. (2013). Multidisciplinary design optimization: A survey of architectures. *AIAA Journal*, *51*(9), 2049-2075. doi: 10.2514/1.J051895

Moore, M. D. (2014). Misconceptions of electric aircraft and their emerging aviation markets. *52nd Aerospace Sciences Meeting*. doi: 10.2514/6.2014-0535

Nicolosi, F., Marciello, V., Cusati, V., & Orefice, F. (2022). Technology roadmap and conceptual design of hybrid and electric configurations in the commuter class. In *Proceedings of the 33rd congress of the international council of the aeronautical sciences.* Stockholm, Sweden: ICAS.

Pornet, C., & Isikveren, A. (2015). Conceptual design of hybrid-electric transport aircraft. *Progress in Aerospace Sciences*, *79*, 114-135. doi: https://doi.org/10.1016/j.paerosci.2015.09.002

Raymer, D. (2018). *Aircraft design: a conceptual approach.* (6th ed.). AIAA Education.

Roskam, J. (2019). *Airplane design part v: Component weight estimation.* Kansas: Roskam Aviation And Engineering Corp.

Schäfer, A. W., Barrett, S. R., Doyme, K., Dray, L. M., Gnadt, A. R., Self, R., ... Torija, A. J. (2019). Technological, economic and environmental prospects of all-electric aircraft. *Nature Energy*, *4*(2), 160–166. doi: 10.1038/s41560-018-0294-x

Torenbeek, E. (2013). *Advanced aircraft design: Conceptual design, technology and optimization of subsonic civil airplanes.* John Wiley Sons.

Zamboni, J., Vos, R., Emeneth, M., & Schneegans, A. (2019). A method for the conceptual design of hybrid electric aircraft. In *Aiaa scitech 2019 forum.* doi: 10.2514/6.2019-1587

Phase Fractions and Velocities in Multiphase Flow – Estimation using Sensor Data Fusion and Machine Learning

Andreas Lund Rasmussen [a], Kjetil Fjalestad [a], Håkon Viumdal [b,*], Ru Yan [b], Saba Mylvaganam [b], Tonni Franke Johansen [c]

[a] *EQUINOR ASA,* [b] *University College of South-Eastern Norway, Faculty of Technology, Natural Sciences and Maritime Sciences, Dept. EE, IT and Cybernetics, Kjølnes Ring 56, 3918 Porsgrunn, Norway,* [c] *SINTEF Digital*
* hakon.viumdal@usn.no

Abstract
There is a strong interest in quantifying the amount of gas and its flow rate to facilitate better control of the processes involved in many industries. There are usually many sensors monitoring these processes, both intrusive and invasive, as well as non-invasive sensors which are usually clamped on to the process pipelines in which the multiphase flow occurs. In the multiphase flow rigs at Equinor and the University of South-Eastern Norway, experiments have been performed with different combinations and velocities of the phases and multiple sensors have been logged. The data from these sensors have been used to estimate volume fractions of the phases as well as their flow rates. This paper presents the estimated results of volume fractions and velocities of selected phases, obtained by fusing data from multiple sensors that monitor density, differential pressure, temperature, and acoustic emission using machine learning (ML) algorithms. These ML algorithms use neural networks with the non-linear input-output type with Levenberg-Marquardt training and provide estimates of volume fractions and phase velocities with RMSE values in the range of 4.6 to 16 m^3/h, with the lowest RMSE for gas and the highest for multiphase flow. The total flow rate for the multiphase flow was in the range 30 to 120 m^3/h. Results are compared with ML models using data from non-invasive sensors.

1. Introduction

It is desirable to know how much oil, water, and gas each well produces in an oil and gas installation. There are many good alternatives to measure single-phase flow with high accuracy, but measuring multiphase flow is more challenging. The sand detectors used on Equinor's oil and gas installations use a piezoelectric sensor to measure the acoustic emission caused by sand production, first reported in a publication of results based on an R&D project supported by STATOIL, the forerunner to the current Equinor, (Folkestad, Mylvaganam, 1990). A multiphase flow does also generate acoustic emission. Since sand detectors are both cheap and non-intrusive Equinor is interested in using their already installed sand detectors to measure acoustic emission from the multiphase flow and fuse this data possibly with other existing measurements for enhancing the process monitoring and control. Together with other available sensors, the process industries are investigating the possibilities of determining the flow rates of individual phases as well the total flow

Related works addressing flow regimes using the multiphase rig have addressed some of the possibilities of identifying flow regimes and flow velocities, , (addressing tomographic approach by Johansen et al, 2018; related to slug modeling and control by Pedersen et al, 2017; modeling single and two-phase flow at 90^o bends in pipelines by Shoux et al 2021; and monitoring flow regime transitions with acoustic and electrostatic sensors in powder flow, Yang et al, 2019). This paper deals with some of the findings from the work done in collaboration with (Rasmussen, A.L., 2023). Some of the related work using the multiphase rig and extensive experiments, but with focus on flow regimes, is presented in the paper in SIMS 2023, (Syed Kazmi et al, 2023).

1.1 System description – Multiphase flow rig with relevant measurands

A dedicated multiphase flow rig has been in use for many years in the Equinor facilities. The system consists of feed pumps for gas, oil and water with the necessary measurements typical for flow related large scale experiments. In the current paper, the focus is only on a sensor suite consisting of the sensors, from which the data used in this study are acquired.

A P&ID of Equinor's multiphase rig is shown in Figure 1. Depending on the phase fractions and flow velocities of the different phases, the production

process can be monitored and controlled using a series of actuators placed near the reservoir (downhole) or at the entrance or exit of the separators.

The volume flow of each medium is indicated by Q_O, Q_W and Q_G with the following equations for Gas Volume Fraction (GVF), Water Liquid Ratio (WLR)

$$GVF = \frac{Q_g}{Q_{tot}}$$

$$WLR = \frac{Q_w}{Q_w + Q_o}$$

with total flow consisting of all the three phases given by

$$Q_{tot} = Q_w + Q_o + Q_g$$

These process variables are of interest in the context of measurement while drilling as well as in process monitoring during the operations of three phase flows in experimental work and oil and gas distribution. These three parameters quantifying the flows and the derived quantities WLR and GVF will be studied using AI/ML techniques with measurements from various sensors, as shown in Figure 1. This study is based on the experiments performed in the Equinor multiphase rig in Herøya, Porsgrunn in the municipality of Telemark.

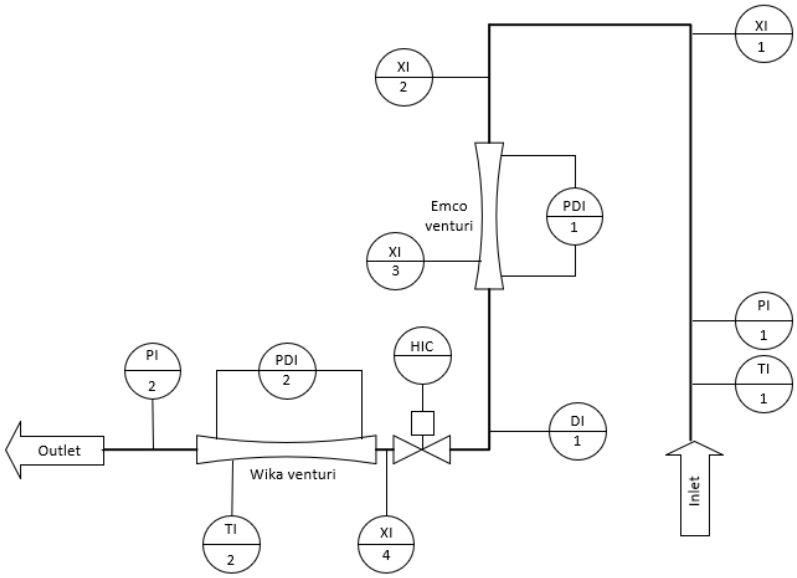


Figure 1. Simplified P&ID (Piping and Instrumentation Diagram) of the Equinor multiphase flow rig with the sensor suite selected for the current study. PDI – Differential Pressure; XI-1, XI-2, XI-3, and XI-4 – Acoustic Emission Sensors (Accelerometers). Inlet can be a single (G, O, W), two-phase (GO, GW, OW), or three-phase flow (O, G, W), with G-Gas, O-Oil, W-Water, with corresponding volume flow rates Q_G, Q_G and Q_W. These variables have been selected as input data for training the artificial neural network models: Density Krohne DI-1, dp 4m straight, Emco Venturi PDI-1, Wika Venturi PDI-2, Accelerometer *i* (XI-i) with *i* =1, ..4, Temperature Difference (TI-1) –(TI-2), Pressure Difference (PDI-1) –(PDI-2).

2. Acoustic emission sensors

The sand detector uses a piezoelectric element to detect the ultrasonic energy from sand particles colliding with the pipe wall and convert it into electrical energy. The sand production can be calculated if the flow velocity is known for the same period. The accelerometers used in the experiments for this paper operate on the same principles as those used in the sand detector. The main difference is whether the sensor detects bubbles/droplets or sand particles. A wiring diagram for a typical sensing system based on acoustic emission is shown in Figure 2.

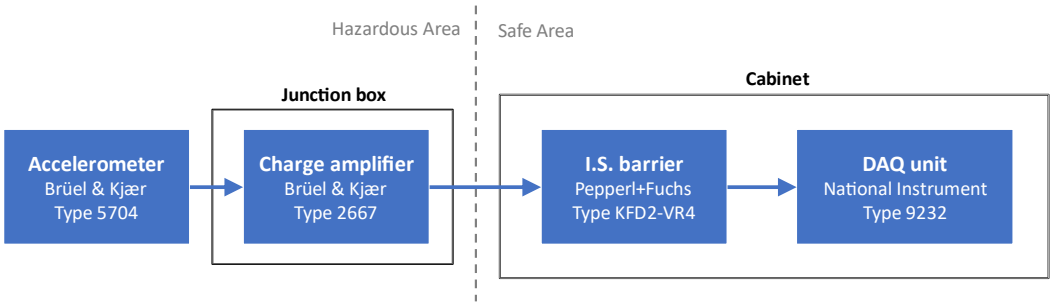


Figure 2. The main modules used in the experiment coupled together in a sensing system based on acoustic emission. I.S. barrier – Intrinsic Safety barrier, **DAQ**- **D**ata **A**c**Q**uisition Unit. The modules in the hazardous area are mounted on the pipe wall or in the vicinity of the multiphase rig, shown with the P&ID in Figure 1.

3. Data Analysis from Experiments

The level of acoustic emission in the multiphase flow varies based on process conditions. When the flow is annular at the entry of the elbow, gas bubbles tend to move to the inner curvature of the bend. Also, the GVF (Gas Volume Fraction) in a gas-liquid flow increases after a restriction in the flowline. Examples of flow restrictions are a Venturi channel or a choke valve.

The experimental data are from 20 distinct experiments involving diverse flow regimes, phase velocities and compositions. In each experiment, the flow rate of the phases are kept constant. The sampling frequency of the accelerometer is 51.2 kHz. Figure 3 presents experimental results involving combined oil and gas flow.

3.1 Analysis of the Root Mean Square (RMS) value of the acoustic emission measurements

The acoustic emission sensors give a voltage signal that can be used for further analysis. The Root Mean Square of voltages, V_{RMS}, of the samples is calculated using equation (1).

$$V_{RMS} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}|V_n|^2} \qquad (1)$$

, where N represents the total number of samples in each experiment, while V_n refers to the voltage of the n^{th} sample.

Figure 3. The letters and numbers in the x-axis contain information about the experiments in dataset 1. The first two letters explain that the experiments are done on gas and oil (GO) flow. The numbers refer to the experiment number, e.g., GO01 – experiment #01 with Gas and Oil, GOC11 – Experiment #11 with Gas and Oil with Choke etc. Legends inserted in each plot indicate the parameters plotted for each experiment. The numbering of the different experiments is the same as in the paper SIMS 2023, (Syed Kazmi et al, 2023).

The V_{RMS} value is based on approximately 2000 samples from each experiment in dataset 1. Each experiment was run under steady-state conditions. The hypothesis is that the V_{RMS} value will correlate with the acoustic emission from the multiphase flow, as was reported in (Folkestad, Mylvaganam, 1990). Pure water flow seems to give low acoustic emission, while pure gas flow seems to give higher acoustic emission. However, pure oil flow seems to give much higher acoustic emission. The observed variations are possible attributed to the differences in the viscosity of the fluids. Significant adjustments in the choke position yield peaks in both total differential pressure and amplitudes of acoustic emission signals. This effect is particularly noticeable for the acoustic emission sensor 4, which is situated downstream of the choke valve, as indicated by the position of the sensor "XI-4" in Figure 1.

Pure water flow Q_w seems to generate low acoustic emission signals, and pure gas flow generates higher acoustic emission signals with gas flow Q_g and total flow Q_{tot}.

3.2 Frequency analysis of the acoustic emission sensors

Figure 4 show the dominant frequency components when increasing the gas flow when the flow contain both gas and oil. Choke valve openings are indicated by percentages in Figure 5, which shows the dominant frequency components in the lower frequency range 3kHz-5 kHz for a completely open choke valve. When the choke valve is gradually closed, there is considerable reduction in the amplitudes of these frequency components. The dominant frequency components seem to be in the range of 0 – 11 kHz for oil and water flow.

4. Preparing the dataset for the artificial neural networks

In addition to the RMS value from the acoustic emission sensors, the selected variables for training the dataset are different density, differential pressure, and temperature measurements. The sampling frequency of the accelerometers is 51.2 kHz, but the other measurements have a sampling frequency of 1 Hz. A code was therefore written to calculate one RMS value every second to get the same sampling frequency for all measurements. The acoustic emission sensors were calibrated for the background noise in the test rig before they were normalized.

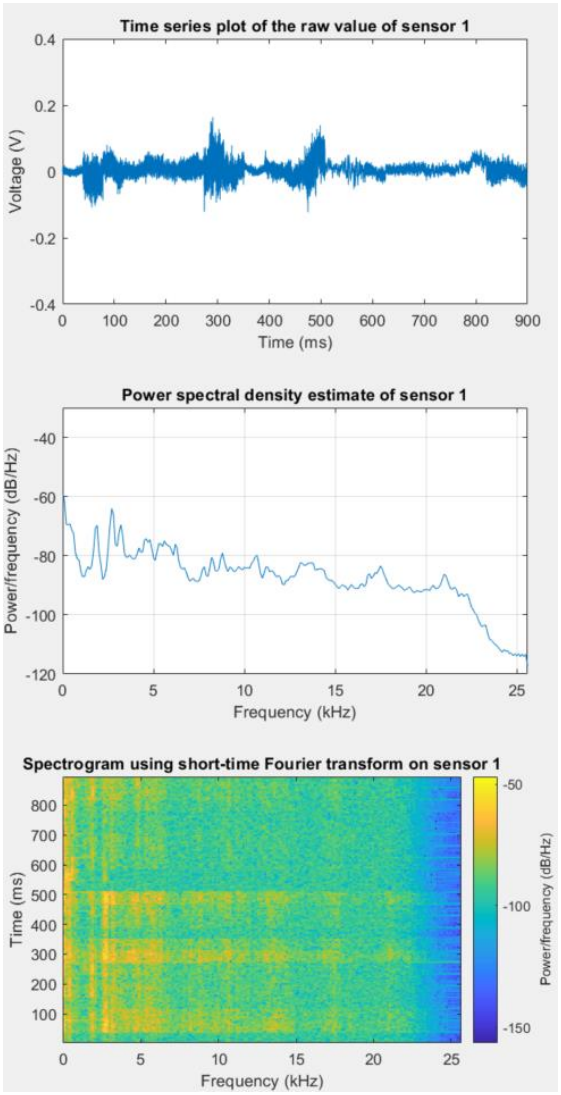


Figure 4. Results based on time series logged from Sensor 1v in GO -flow, as indicated by the position of "XI-1" in Figure 1. Spectrogram and power spectral density show incidents of increased flow in the pipe.

Both linear scaling and Z-score normalization methods were used, but Z-score gave overall best results when testing the models on other datasets. The Z-score normalization method assumes no extreme outliers is given in equation (2).

$$x' = \frac{(x-\mu)}{\sigma} \tag{2}$$

, with x representing the sample value, μ the ensemble average and σ the standard deviation.

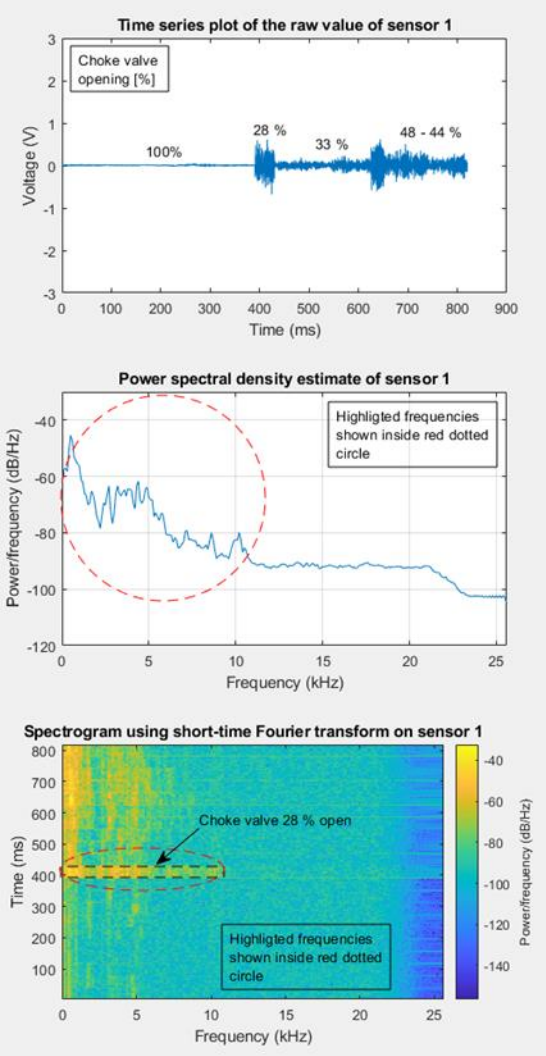


Figure 5. Results from time series logged in from Sensor 1, in OW-flow, as indicated by the position of "XI-1" in Figure 1. Both power spectral density plot and spectrogram show the effects of opening the choke valve as indicated by the circle and ellipse shown with the red dashed lines.

5. Neural network for multiphase flow estimation

The shallow nonlinear input-output network using Z-score normalization gave the best results. An example of a nonlinear input-output model is shown in Figure 6. The most important results use all four acoustic emission sensors "XI" and the differential pressure measurement over Venturi 2 "PDI-2" as inputs to the network. The sensors are all shown in the P&ID in Figure 1. There is one network each for gas, oil, water, and total flow rate. The GVF, and LWR (Water Liquid Ratio) are calculated from the flow rates.

6. Results

To illustrate the performance of the model, one set of plots comparing the actual and estimated parameters is shown in Figure 7.

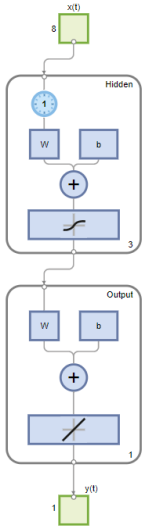


Figure 6. Nonlinear input-output neural network with one hidden layer and one output layer. This network is configured with a time delay of 1, 3 neurons in the hidden layer and 1 output in the output layer. Inputs, with reference to Figure 1: Differential pressures on two locations, acoustic emission signals (XI-I, i=1,2,3,4), density from DI-1, Difference of Temperatures TI1 and TI2. Output: gas flow rate.

7. Discussion

The frequency analysis indicates that pure oil flow gives a higher amount of acoustic emission than pure gas flow and that pure water flow gives almost zero acoustic emission. It is recommended to investigate this finding in future research since there may be a relationship between with the viscosity of the fluid, or some other variable that can be exploited. It seemed like the frequency range between 0 – 11 kHz contained the most important information. Since there was no clear conclusion in the frequency analysis all frequency components were kept avoiding the risk of filtering away important information. This decision may have been a mistake. Further investigation is needed on the frequency analysis to find out if some of the frequency components are due to noise, but also to research if there are some dominant frequency components in gas, oil, or water flow. If there is a dominant frequency component in, for instance, gas, then the amplitude of that frequency component may correlate with the gas flow rate. In the frequency analysis it was shown that there may be a relationship between frequency components around 1.7 kHz and total and water flow, but this needs to be investigated further.

The models for gas, oil, water, and total flow were trained with shallow neural networks. It is recommended to use a deep neural network in future studies to improve the results. According to MATLAB documentation, the NARX model will give more accurate predictions than the nonlinear input-output models since it uses the additional information from previous values of $y(t)$.

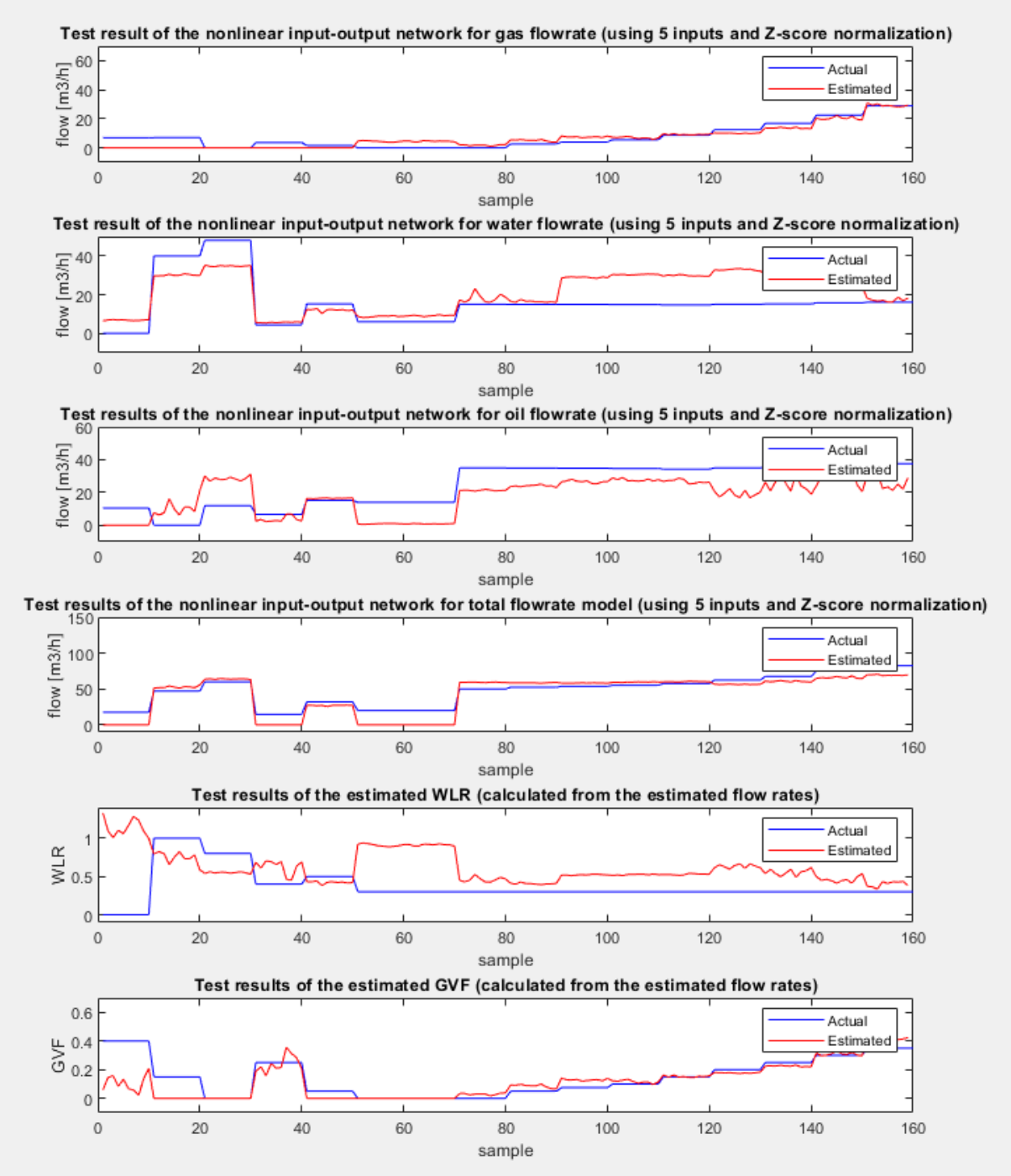


Figure 7. Comparison of estimated and actual results for WLR (Water Liquid Ratio) and GVF (Gas Volume Fraction). The 5-inputs: are: datafrom the four acoustic emission sensors (XI-I, i=1,2,3,4) and differential pressure PDI-2 over Wika Venturi meter, based on the P&ID shown in Figure 1.

In this study, it was observed that the NARX models gave the best training results, but that after closing the loop and simulating the network on the testing matrix the results were not as good as expected. The Nonlinear Input-Output models gave the best results when using the testing matrix. This network does not use the previous values of $y(t)$ as feedback as the NARX network does. This may indicate that there is no relationship between the measured flow rate and the previously measured flow rate. The models using

7-8 inputs, gave the best results. The models with only 5 inputs, were used since they gave similar results. The models with 5 inputs are recommended since they are a more cost-effective solution with respect to execution time and resources.

There seems to be a direct relationship between the acoustic emission generated by the multiphase flow and the differential pressure, which again correlates with the total flow rate. The Root Mean Square Error (RMSE) values in Table 1 reveal that the RMSE for oil and water flow rate models are higher than for the gas flow rate model, implying less accuracy in the oil and water models which will need improvement. Since the total flow rate is the sum of oil, water, and gas flow rate, its RMSE value is higher, but the results are satisfactory. The GVF and WLR are calculated from flow rates and are therefore dependent on the accuracy of the flow rate models.

The location of the acoustic emission sensors may be vital for the results. It was shown in other studies that the GVF is increased downstream of a flow restriction like a Venturi or choke valve. Note that this is only true in the pipe just after the flow restriction and not for the whole pipe downstream the flow restriction. If this is to be interpreted as something that can be exploited in future studies or if this should be conceived as measurement noise should be investigated. If a sand detector is chosen as the acoustic emission sensor, then the location may not be ideal since it is located to best detect sand particles that collide with the outer curvature of the pipe wall, while gas bubbles seem to flow closer to the inner curvature of a 90-degree bend.

8. Conclusion

We implemented shallow nonlinear input-output neural networks for gas, oil, water, and total flows. Although the inclusion of additional sensors marginally improved accuracy, the cost-effective solution using only five measurements as input was favored due to the insignificant difference.

Equinor is keen on utilizing already installed huge number of sand detectors to measure flow velocities based on acoustic emission caused by the multiphase flow. This study suggests the potential of combining four acoustic emission sensors with differential pressure measurements over a Venturi meter to estimate multiphase flow velocities.

Shallow nonlinear input-output neural networks were created for gas, oil, water, and total flow. While including more sensors did improve the accuracy, the difference was not substantial, leading to a preference for the more cost-effective solution that only required five input measurements.

In future studies, it is recommended to use a deep neural network to improve the results to improve the results, and place greater emphasis on frequency analysis to identify dominant frequency components. The most interesting frequency range in this study appeared to be 0 – 11 kHz; however, due to the lack of a clear conclusion, all frequency components were retained throughout the study.

Table 1. An overview of the RMSE, network algorithm and configuration. The models for oil and water flow have room for improvement, while the gas and total flow models have both good accuracy and low RMSE. Dataset 1 was only available in, but dataset 2 was preferred to be used for training the data. This was due to a higher sampling rate on measurements and more available sensors.

	Network type	Training algorithm	Network configuration	RMSE
Gas flow model	Nonlinear Input-Output model	Levenberg-Marquardt	1 hidden layer with 2 neurons	4.62
Oil flow model	Nonlinear Input-Output model	Levenberg-Marquardt	1 hidden layer with 2 neurons	11.40
Water flow model	Nonlinear Input-Output model	Levenberg-Marquardt	1 hidden layer with 2 neurons	10.87
Total flow model	Nonlinear Input-Output model	Levenberg-Marquardt	1 hidden layer with 2 neurons	16.65

Acknowledgement

Many thanks to the team working with Multiphase Flow in Equinor for performing dedicated measurement campaigns to acquire the necessary data used in this study. This work was done as part of the activities of the project, SAM (SAM: Self Adapting Model-based system for Process Autonomy - SINTEF), with Equinor and SINTEF as the major partners.

References

Folkestad, T., Mylvaganam, K.S., 1990, Acoustic measurements detect sand in North Sea flow lines. Oil and Gas Journal; (USA). 88:35

Johansen, R., Østby, T.G., Dupre, A., Mylvaganam, S. , 2018, Long short-term memory neural networks for flow. 9th World Congress on Industrial Process Tomography, Bath, UK,

Pedersen, S., Durdevic, P., Yang, Z., 2017, Challenges in slug modelling and control for offshore oil and gas, International Journal of Multiphase Flow Vol. 88, pp. 270-284,

Rasmussen, A.L., 2022, Multiphase flow metering – estimation of flow velocities of component phases using multimodal sensor suite, Master Thesis, University of South -Eastern Norway

Shouxu Q., Wenyi Z., Shuang, W., Lanxin, S., Sichao, T. 2021, Numerical simulation of single and two-phase flow across 90° vertical elbows, Chemical Engineering Science, Volume 230, 116185, ISSN 0009-2509, https://doi.org/10.1016/j.ces.2020.116185.

Syed Kazmi, N., Yan, R., Mylvaganam, M., Viumdal, H., 2023, "Multimodal sensor suite for identification of flow regimes and estimation of phase fractions and velocities – Machine learning algorithms in multiphase flow metering and control, SIMS 2023

Wang,x.,Liu,Y., Xin,H.,2021,Bond strength prediction of concrete-encased steel structures using hybrid machine learning methodructures,Volume 32,Pages 2279-2292,ISSN 2352-0124,https://doi.org/10.1016/j.istruc.2021.04.018.

Yang Y., et al, 2019, Measurement and analysis of flow regimes transition by acoustic and electrostatic signals in vertical pneumatic conveying, Powder Technology, Volume 352, 2019, Pages 283-293, ISSN 0032-5910,https://doi.org/10.1016/j.powtec.2019.04.024.

Process Simulation and Cost Optimization of a Gas based Power Plant including amine-based CO_2 Capture

Lars Erik Øi, Esmaeil Aboukazempour

Department of Process, Energy and Environmental Technology, University of South-Eastern Norway

Lars.oi@usn.no

Abstract

CO_2 capture from gas turbine exhaust gas using heat from the power generation cycle is a possibility for CO_2 emission reduction from natural gas-based power plants. A simplified power plant was simulated in Aspen HYSYS with a compressor, a combustion chamber, a turbine, a steam circuit with a steam heater, a high-pressure steam turbine, a low-pressure steam turbine, a steam condenser, and a circulating pump. CO_2 capture was simulated with an absorption column, a rich amine pump, a lean/rich amine heat exchanger, a desorber with a reboiler and condenser, a lean pump and an amine cooler. The equipment cost was obtained from Aspen In-plant Cost Estimator, and an enhanced detailed factor method was used to estimate the total investment. A base case with combustion at 30 bar, ΔT_{MIN} of 10 °C, and 10 stages (meters of absorber packing) was simulated, dimensioned, and cost estimated. In earlier works, optimum parameters have been found by minimizing the cost of CO_2 capture. In this work, optimum was defined as the maximum profit for a combined process with 85 % capture efficiency. Optimized parameters were calculated to 25 bar for the combustion pressure, 13 °C for the minimum temperature approach in the lean/rich amine heat exchanger, and 10-meter packing height in the absorption column. These values are comparable to values in literature.

Keywords: *Carbon capture, Aspen HYSYS, gas turbine, cost estimation, simulation.*

1. Introduction

CO_2 capture from fossil fuel-based power plants is extensively studied in the literature (Liang et. al, 2015; Li et al., 2016). Some papers include power production and CO_2 capture in their studies (de Ruick, 1992; Kvamsdal et al.., 2007; Øi, 2007; Mathisen et al., 2011; Schach et al., 2010; Amrollahi, 2012; Karimi et al., 2012; Hu et al., 2017; Luo et al., 2015). There is also much literature on the cost estimation of CO_2 capture (Rao and Rubin, 2002; Ali, 2019; Shirdel et al., 2022).

Research work on the combination of simulation, cost estimation and cost optimization of CO_2 capture has been performed by Kallevik (2010), Øi (2012) and Shirdel et al. (2022). In a Ph.D. Thesis by Ali (2019), the Enhanced Detailed Factor (EDF) method was presented. Øi et al. (2021) evaluated the automated calculation of cost optimum process parameters in the CO_2 capture process. Typical parameters to optimize in an amine-based process are the number of stages in the absorption column and the minimum temperature approach in the heat exchangers.

Nord et al. (2017), Kazemi et al. (2022) and Øi et al. (2022) have evaluated simulation and cost evaluation of combined power generation and CO_2 capture. Nord et al. (2017) and Øi et al (2022) evaluated this for an offshore application.

For a natural gas-based power plant, typically a Natural gas combined cycle (NGCC) plant, the tools GTPRO, GTMASTER and GateCycle have been used for simulation. For CO_2 capture, the programs Aspen Plus, Aspen HYSYS and Unisim are standard programs. One paper (He and Ricardez-Sandoval, 2016) has used Aspen Plus for both an NGCC and a CO_2 capture process, and one paper (Hu et al., 2017) has used Unisim for both parts.

This study presents findings derived from the master's thesis by Aboukazempour (2023). The focal point of this research lies in the simulation and cost optimization of a natural gas-based power plant with integrated CO_2 capture using Aspen HYSYS simulation software. A novel aspect of this investigation involves the integration of a power plant and a carbon capture facility which enable the cost optimization of several key parameters, including the power plant's inlet pressure, the number of stages in the absorption column and the minimum temperature difference in the heat exchangers. These optimum parameter values were found by optimization of the net present value assessment of the entire system.

2. Process Description and Specifications

2.1. Process Description

Fig. 1 shows a simplified NGCC process. Natural gas is combined with compressed air in the combustion chamber and produces power in the expander part of the gas turbine. The turbine exhaust heats steam in the steam generator which produces power in a steam turbine before it is pumped back to the steam generator.

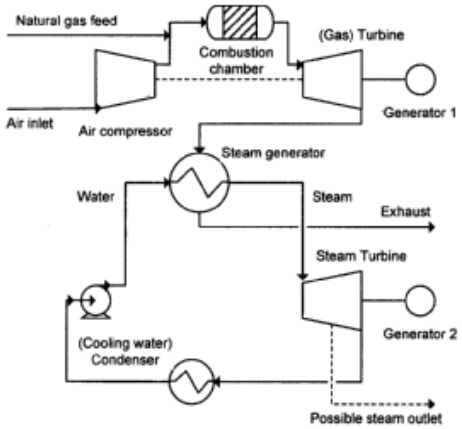


Figure 1: Combined cycle power plant (Øi, 2007)

Fig. 2 shows the CO_2 capture process. The flue gas is cooled in the Direct Contact Cooler (DCC) before the CO_2 is absorbed in lean amine in the absorber. The rich amine from the bottom is pumped through a heat exchanger to the desorber where CO_2 is the top product and regenerated lean amine from the bottom is returned through the heat exchanger and a cooler to the absorption column.

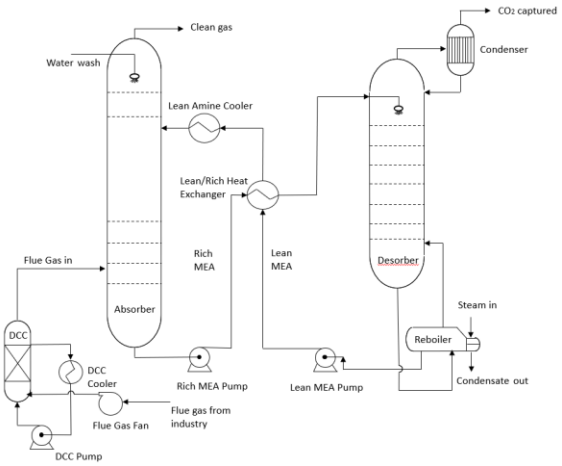


Figure 2: Process flow diagram of a standard amine-based CO2 capture process (Aromada et al., 2020)

2.2. Process Specifications and Simulation

The process specifications used for the base case simulation are presented in Tables 1 and 2. The process simulation in this work is similar to the work of Øi (2007). The simulations were performed in Aspen HYSYS Version 12. The base case was simulated to capture 85 % CO_2 from exhaust gas from the simplified NGCC power plant. The process has a 10 ℃ temperature difference in the main heat exchanger.

The Aspen HYSYS process flow diagram showing all the equipment included in the scope of the study is shown in Fig. 3.

Table 1: Aspen HYSYS specifications for base case power plant model

Parameter	Value
Inlet air temperatures	25 ºC
Inlet natural gas pressure	30 bar
Combustion temperature	1500 ºC
Steam high pressure	120 bar
Steam medium pressure	3.5 bar
Steam low pressure	0.07 bar
Pressure to stack	1.01 bar
Stack temperature	100 ºC

Table 2: Specifications for the CO_2 capture process

Items	Specifications [Unit]	Value
Inlet Flue Gas	Temperature [°C]	40
	Pressure [bar]	1.1
	Molar flow rate [kmol/h]	71345
	CO_2 content [mole %]	4.61
	H_2O content [mole %]	6.71
Lean MEA	Temperature [°C]	40
	Pressure [bar]	1.1
	Molar flow rate [kmol/h]	99496
	MEA content [W %]	28.92
	CO_2 content [W %]	5.39
Absorber	Number of stages	10
	Murphree efficiency	0.25
	Rich amine pump pressure [bar]	2
	Rich amine temp. out of Lean/Rich amine HEx [°C]	102.7
Desorber	Number of stages in stripper	6
	Murphree efficiency	1.00
	Reflux ratio in the desorber	0.3
	Reboiler temperature [°C]	120
	Pressure [bar]	2
	Lean amine pump pressure [bar]	5

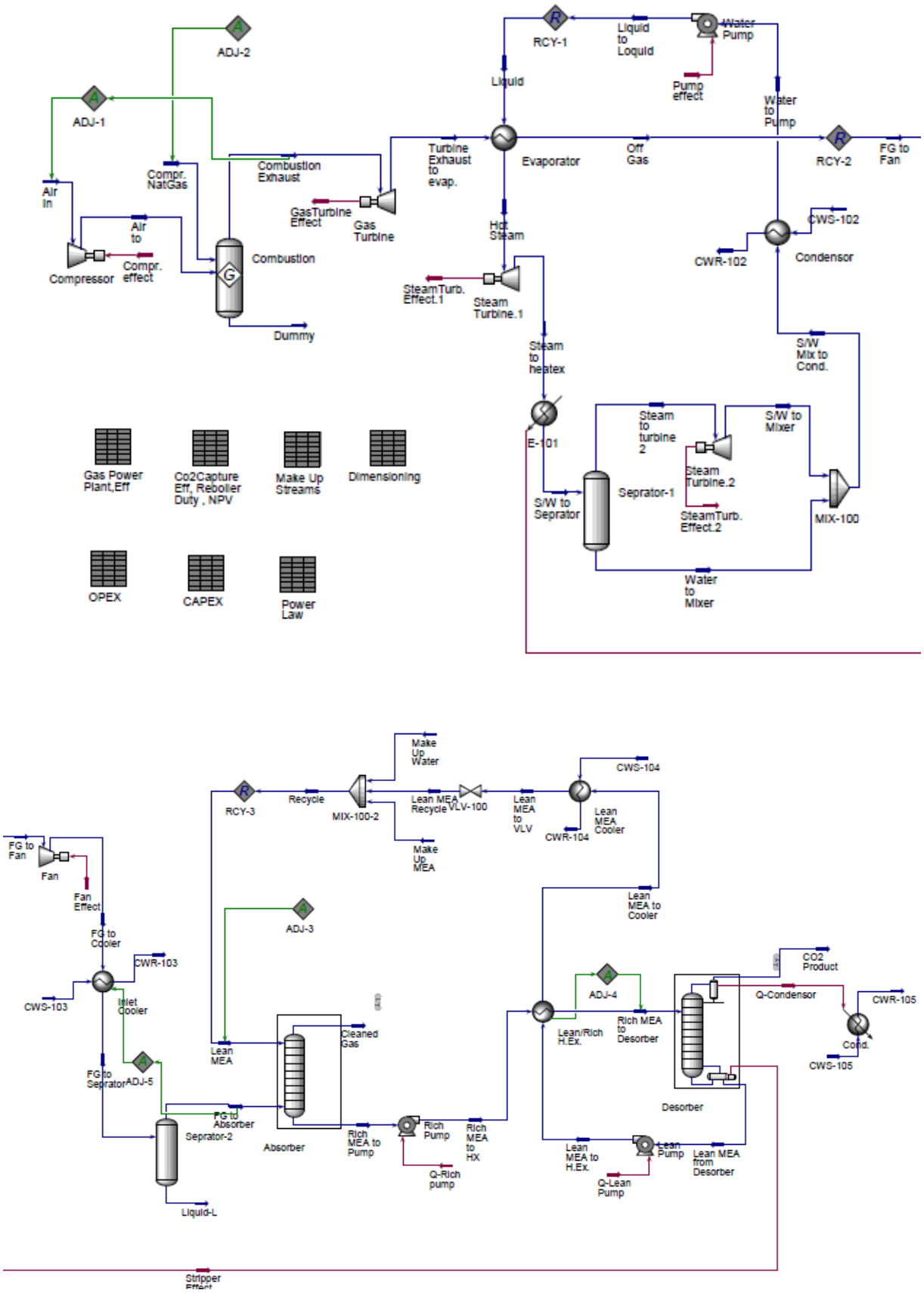


Figure 3: Combined power plant and CO_2 capture process flowsheet in Aspen HYSYS (Aboukazempour, 2023)

2.3. Equipment Sizing

Murphree efficiencies of 0.25 and 1.00 were specified for all the absorber and desorber stages, respectively. For the absorber and desorber internals, structured packing was assumed, and one stage was assumed to correspond to 1 meter of packing height. The column diameters were calculated based on a gas velocity of 2.5 m/s and 1 m/s, respectively, as in Park and Øi (2017) and Øi et al. (2021). The total height of the absorption column and desorption column were specified to be packing height plus 25 m and 15 m respectively (Kallevik, 2021) due to distributors, water wash packing, demister, gas inlet, gas outlet, and sump. 75 % adiabatic efficiency was specified in the pumps, compressors and expanders. Overall heat transfer coefficient values have been specified for the lean/rich heat exchanger to 732 W/(m^{2}K). These values are close to the same as in Øi (2012) and Park and Øi (2017) and slightly less than the numbers in Øi et al. (2021) which are regarded as optimistic.

2.4. Capital and Operating Cost Estimation

The equipment costs were calculated in Aspen In-plant Cost Estimator version 12, which gives the cost in Euro (€) for the Year 2020. A generic location (e.g. Rotterdam) was assumed and stainless steel (SS316) with a material factor of 1.3 was assumed for all equipment units. In a detailed factor method like the EDF method, each equipment cost (in carbon steel) was multiplied by an installation factor to get the equipment installed cost. The installation factor is a function of the site, equipment type, materials, and size of the 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 2020 (Aromada, 2021) were used. The specifications for operating cost estimation are found in Table 3.

Table 3: Cost calculation specifications.

Item	Value	Unit
Operating Lifetime	25	[year]
Construction Lifetime	3	[year]
Operation Lifetime	22	[year]
Discount rate	7.5 %	-
Operating Hours	8000	[h/year]
Electricity Price	0.136	[€/kWh]
Natural gas Price	1.29	[€/m3]
Cooling water Price	0.022	[€/m3]
Water process Price	0.203	[€/m3]
Solvent MEA Price	1450	[€/ton]
Maintenance Price	4% of CAPEX	[€/year]
Operator Price	80414 × (12 Operators)	[€/year]
Engineer Price	156650 × (2 Engineer)	[€/year]

2.5. Net Present Value (NPV) and Payback Period

Cost optimization can be based on the maximization of the net present value (NPV) of the project. This common measure is defined by Equation (1) for a defined process plant and a defined time of operation.

$$NPV = CAPEX + NPV_{OPEX} \quad (1)$$

Where:

- NPV= Net present value for the total costs [€]
- CAPEX = Installation expenses for equipment [€]
- NPV_{OPEX} = The total cost of OPEX for the calculation period [€]

In this work, the NPV_{OPEX} cost for the calculation period is calculated and added to the CAPEX cost to obtain the total NPV. The NPV_{OPEX} cost is obtained by Equation 2:

$$NPV_{OPEX} = \sum_{N=3}^{End} \{(a) \times \frac{1}{(1+i)^N}\} \quad (2)$$

Where:

- i = annual interest rate
- a = annual operation cost [€]
- N = number of years

The calculated NPV in Equation (1) consider all the incomes and costs related to the utilities and the CAPEX. The NPV of the early years is negative, but it will be positive after this period due to the income related to the sale of electricity. A higher NPV indicates that the project is more profitable. The cost calculation shows a €1570 million net present value over a 25-year plant lifetime.
As seen in Fig. 4, after six years of operation after construction (the ninth year in the table), the NPV of the project becomes positive, which indicates a six-year payback period.

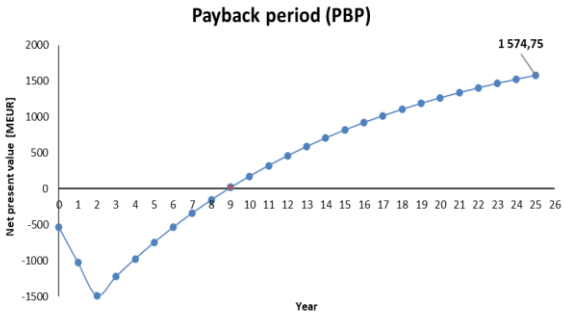


Figure 4: Payback time for the base case (Aboukazempour, 2023)

3. Results and Discussion

3.1. Simulation Results of the Base Case Model

The 400 MW net electricity output and 85% CO_2 removal are the main adjusted parameters in this Aspen HYSYS simulation. Based on these goals these three main parameters inlet pressure into the power plant, the number of stages in the absorption column, and the minimum temperature approach in the lean/rich heat exchanger were optimized based on the net present value during the total lifetime project.

3.2 Optimization of Combustion Pressure

Fig. 5 shows NPV as a function of inlet pressure and combustion pressure with an optimum of 25 bar. Earlier suggestions for optimum pressure vary between 15 and 35 bar. 18 bar has been suggested as an optimum by Horlok (2003), but he claims that 30 bar is optimum for a gas turbine operating alone. Soares (2015) states that a higher combustion temperature favors a higher pressure with 12 bar at 1100 K and higher than 40 bar at 1800 K. Ibrahim et al (2011) state that the highest total efficiency of combined cycle gas turbines takes place at a high compression pressure ratio with low ambient temperature as in this work. The gas turbine in this work operates at about 1800 K. In the modern heavy-duty Siemens gas turbine, the pressure is 24 bar which is close to the optimum for this work. The optimization in this work is a simplification but shows that a combined model gives reasonable results compared to earlier optimization by e.g. Horlok (2003) and Ibrahim et al. (2011).

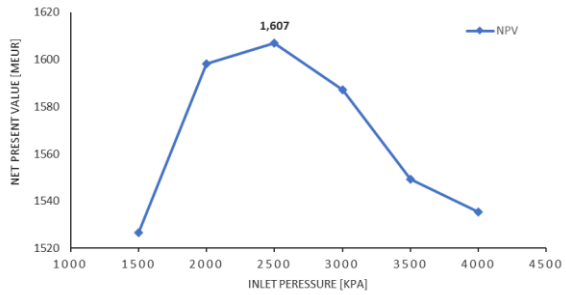


Figure 5: NPV calculation as a function of inlet pressure for base case (Aboukazempour, 2023)

3.3 Optimization of Minimum Temperature Approach in lean/rich amine heat exchanger

Fig. 6 shows NPV as a function of the minimum temperature approach in the main lean/rich amine heat exchanger. The figure shows a flat optimum between 13 and 17 °C with an optimum at 13 °C. There are several sources indicating an optimum between 10 and 15 °C (Øi, 2012; Øi et al., 2021; Aromada et al., 2022; Shirdel et al., 2022; Øi et al., 2022). It is well-known that the optimum is quite flat.

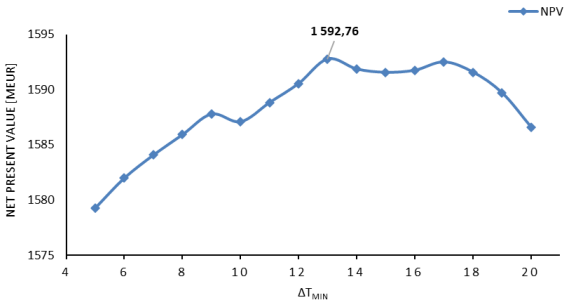


Figure 6: Manual NPV calculation results as a function of the minimum temperature approach (Aboukazempour, 2023)

Fig. 7 shows NPV as a function of the minimum temperature approach for both manual and automated calculations. The manual calculations give a smoother curve. The reason is that the manual calculations (or manual adjustment of the convergence) can adjust the convergence more accurately. The automated calculations are based on a case study, where some of the recycles are not adjusted. The optimum temperature approach is 13 °C for the manual calculation and 14 °C for the automated calculation.

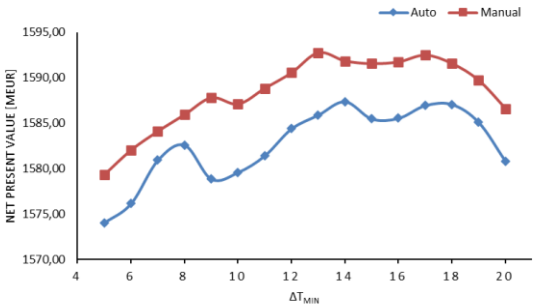


Figure 7: Comparison of manual and automatic NPV results for the minimum temperature approach. (Set Point: ΔTmin = 10 °C), (Aboukazempour, 2023)

3.4 Optimization of the number of absorption stages

Fig. 8 depicts the relationship between NPV and the number of stages in the absorber column, showing an optimal configuration at 10 stages (measured in meters of packing). In contrast, alternative references have computed optimal column heights within the higher range of 12 to 18 meters, as in works like Amrollahi (2011), Øi (2012), Aromada et al. (2022), Shirdel et al. (2022), Øi et al. (2021) and Øi et al. (2022).

The greater number of stages observed in preceding absorber studies could be attributed to the pursuit of more ambitious CO_2 removal efficiency targets. Within the scope of this research, achieving an 85% CO_2 removal rate would likely correspond to a lower optimal absorption column height.

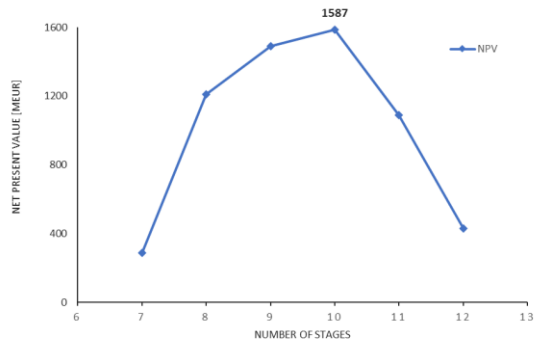


Figure 8: Manual NPV calculation results for the number of stages in the absorber column (Aboukazempour, 2023)

3.5 Modified Base Case Model

The initial scenario was modified using the best parameters identified in the sensitivity analysis. Table 4 displays the optimal process parameters employed in this adjusted scenario simulation.

Table 4: Aspen HYSYS optimum parameters results based on the net present value (NPV), (Aboukazempour, 2023)

Modified parameter	Value
Inlet pressure into the power plant	2500 kPa
Minimum approach temperature (ΔT_{min})	13 °C
Number of stages in the absorption column	10

The modified scenario projected a net present value of 1900 million Euro compared to 1600 million Euro for the initial scenario. The payback duration was reduced from 6 years to 5 years. Fig. 9 illustrates the payback period of the modified scenario based on the net present value of the project.

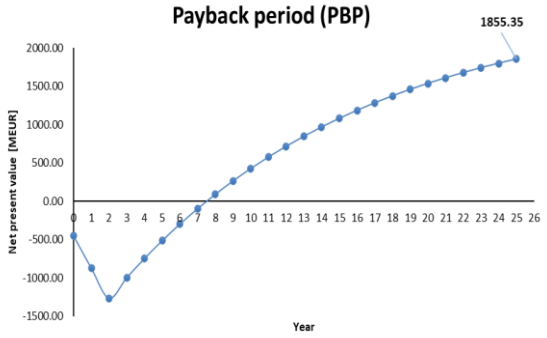


Figure 9: Payback period of the modified base case. (Aboukazempour, 2023)

3.6 Simultaneous parameter optimization

While the current focus in this paper is on refining individual parameters, it is essential to recognize the potential benefits of optimizing all relevant factors simultaneously. This was performed by Karzemi et al. (2022), but this was only energy optimization and not economic optimization. Although it is concentrated on optimizing combustion pressure, absorption column stages and heat exchanger temperature in this study, it is worth noting that these parameters are interconnected. A comprehensive optimization approach may lead to more accurate and refined optimum results.

Although simultaneous optimization of several parameters is not calculated in this study, Aspen HYSYS provides a platform for such investigations.

Apart from Aspen HYSYS, other software tools like Aspen Plus, Unisim, and GateCycle also offer capabilities for simultaneous optimization of several parameters.

3.7 Accuracy, uncertainties and limitations

Uncertainties from process assumptions, cost estimation and parameter adjustments significantly impact the study's precision. Notably, there are substantial uncertainties in estimating process equipment costs, especially regarding main equipment installation expenses. This concern is particularly pronounced for high-cost components such as compressors and gas turbines. The cost of natural gas also has a considerable influence on the net present value of a power plant with or without CO_2 capture.

4. Conclusion

A standard gas-based power plant process including CO_2 capture based on absorption into monoethanolamine (MEA) has been simulated and cost estimated with an equilibrium-based model in Aspen HYSYS2.0. The power plant exhaust is the input to the CO_2 capture simulation, and the steam demand for CO_2 capture is the input to the power plant simulation.

The power plant was calculated with a compressor, a combustion chamber, a turbine, a steam circuit with a steam heater, a high-pressure steam turbine, a low-pressure steam turbine, a steam condenser, and a circulating pump. The CO_2 capture plant was simulated with an absorption column, a rich amine pump, a lean/rich amine heat exchanger, a desorber with a reboiler and condenser, a lean pump, and an amine cooler. The equipment cost was obtained from Aspen In-plant Cost Estimator V12.0, and an enhanced detailed factor (EDF) method was used to estimate the total investment. A base case with combustion at 30 bar, ΔT_{MIN} of 10 °C and 10 stages (meters of packing in the absorber) was simulated, dimensioned and cost estimated.

In earlier works, cost optimum parameters have been found by minimizing the cost of CO_2 capture. In this work, optimum was defined as the maximum profit for a combined process with 85 % capture efficiency.

Optimized parameters were the minimum temperature approach in the lean/rich amine heat exchanger, the number of absorption column stages and the combustion pressure in the power plant. The optimum values were calculated to 13 °C, 10 stages, and 25 bar. These values are comparable to values in the literature.

References

Aboukazempour, E. (2023), *Process Simulation and Cost Optimization of a Gas based Power Plant including amine-based CO_2 capture*. Master's Thesis, University of South-Eastern Norway, Porsgrunn.

Ali, H. (2019) 'Techno-economic analysis of CO_2 capture concepts', PhD Thesis, University of South-Eastern Norway. [Online]. Available: http://hdl.handle.net/11250/2622802

Amrollahi, Z. (2012) 'Optimized process configurations of post-combustion CO_2 capture for natural-gas-fired power plant – Power plant efficiency analysis', *Int. J. Greenh. Gas Control*, vol. 8, p. 1-11, doi: 10.1016/j.ijggc.2012.01.005.

Aromada, S. A. et al. (2021) 'Capital cost estimation of CO_2 capture plant using Enhanced Detailed Factor (EDF) method: Installation factors and plant construction characteristic factors', *Int. J. Greenh. Gas Control*, vol. 110, no. 1, p. 103394, doi: 10.1016/j.ijggc.2021.103394.

De Ruyck, J. (1992) 'Efficient CO_2 capture through a combined steam and CO_2 gas turbine cycle', *Energy Convers. Manag.*, vol. 33, no. 5–8, pp. 397–403, [Online]. Available: https://www.sciencedirect.com/science/article/pii/019689049290036V

He, Z. and Ricardez-Sandoval (2016) 'Dynamic modelling of comercial-scale CO_2 capture plant integrated with a natural gas comined cycle (NGCC) power plant', *Int. J. Greenh. Gas Control*, vol. 55, pp. 23-35, doi: 10.1016/j.ijggc.2016.11.001.

Horlok, J.H. (2003) 'Advanced gas turbine cycles', Pergamon Press.

Hu, G. et al. (2017) 'Thermodynamic analysis and techno-economic evaluation of an integrated natural gas combined cycle (NGCC) power plant with post-combustion CO_2 capture and heat supply', *Applied Thermal Engineering*, 111, pp. 308-316, doi: 10.1016/j.applthermaleng.2016.09.094.

Ibrahim, T.K. and Rahman, M.M. (2012) 'Effect of Compression Ratio on Performance of Combined Cycle Gas Turbine', *IJEE*, 2(1), pp. 9-14, doi: 10.5923/j.ijee.20120201.02.

Kallevik, O. B. (2010) 'Cost estimation of CO_2 removal in HYSYS.' Master's thesis. Telemark University College, Norway.

Karimi, M. et al. (2012) 'Natural Gas Combined Cycle Power Plant Integrated to Capture Plant', Energy & Fuels, 26(3), pp. 1805-1813, doi: 101021/ef201921s.

Kazemi, A., Moreno, J., & Iribarren, D. (2022) 'Techno-economic comparison of optimized natural gas combined cycle power plants with CO2 capture', Energy, vol. 255, no. 124617, doi: 10.1016/j.energy.2022.124617

Kvamsdal, H. M. et al. (2007) 'A quantitative comparison of gas turbine cycles with CO_2 capture', *Energy*, vol. 32, no. 1, pp. 10–24, doi: 10.1016/j.energy.2006.02.006.

Li, K. et al. (2016) 'Systematic study of aqueous monoethanolamine (MEA)-based CO_2 capture process: Techno-economic assessment of the MEA process and its improvements. *Applied Energy*, 165, 648-659.

Liang, Z. et al. (2015) Recent progress and new developments in post-combustion carbon-capture technology with amine-based solvents. *International Journal of Greenhouse Gas Control*, 40, 26-54. doi: 10.1016/j.ijggc.2015.06.017.

Luo, X and Wang, M. (2015) 'Optimal operation of MEA-based post-combustion carbon capture for natural gas combined cycle power plants under different market conditions," *International Journal of Greenhouse Gas Control*, 48, pp. 312-320. doi: 10.1016/j.ijggc.2015.11.014.

Mathisen, A. et al. (2011) 'Combining bioenergy and CO_2 capture from gas fired power plant' *Energy Procedia*, 4: pp. 2918-2925, doi. 10.1016/ef201921s.

Nord, L. O. R. et al. (2017) 'CCS on Offshore Oil and Gas Installation - Design of Post-Combustion Capture System and Steam Cycle', *Energy Procedia*, vol. 114, no. 7, pp. 6650–6659, Jul. 2017, doi: 10.1016/j.egypro.03.1819.

Park, K. and Øi, L. E. (2017) 'Optimization of gas velocity and pressure drop in CO_2 absorption column', Linköping Electronic Conference Proceedings SIMS 58, pp. 292-297. doi: 10.3384/ecp17138292

Rao, A. B., and Rubin, E. S. (2002) 'A Technical, Economic, and Environmental Assessment of Amine-Based CO_2 Capture Technology for Power Plant Greenhouse Gas Control', *Environ. Sci. Technol.*, vol. 36, no. 20, pp. 4467–4475, doi: 10.1021/es0158861.

Schach, M. O. et al. (2010) 'Techno-economic analysis of post combustion processes for the capture of carbon dioxide from power plant flue gas', *Industrial & Engineering Chemistry Research*, 49(5), 2363-2370,

Shirdel S. *et al.*, (2022) 'Sensitivity Analysis and Cost Estimation of a CO_2 Capture Plant in Aspen HYSYS', *ChemEngineering*, vol. 6, no. 2, p. 28, doi: 10.3390/chemengineering6020028.

Soares, C. (2015) 'Chapter 19 – Basic Design Theory' in Gas Turbines (2nd ed.), Butterworth-Heinemann.

Øi, L. E. (2007) 'Aspen HYSYS simulation of CO_2 removal by amine absorption from a gas based power plant. In The 48th Scandinavian Conference on Simulation and Modeling (SIMS 2007), *Linköping University Electronic Press*, 73-81.

Øi, L. E. (2012) 'Removal of CO_2 from Exhaust Gas', Ph.D. Thesis, Department of Process, Energy and Environmental Technology, Telemark University College, Porsgrunn, Norway.

Øi, L. E. *et al.* (2021) 'Automated Cost Optimization of CO_2 Capture Using Aspen HYSYS', Linköping Electronic Conference Proceedings SIMS 62, pp. 293-300. doi: 10.3384/ecp21185293

Øi, L. E. *et al.* (2022) 'Process simulation and automated Cost Optimization of CO_2 Capture Using Aspen HYSYS', Linköping Electronic Conference Proceedings SIMS 63, pp. 65-71.

ESP Lifted Oil Field: Core Model, and Comparison of Simulation Tools

Bernt Lie[a,*]

[a] *University of South-Eastern Norway, Porsgrunn, Norway*
[*] Bernt.Lie@usn.no

Abstract

Optimal operation of petroleum production is important in a transition from energy systems based on fossil fuel to sustainable systems. One sub-process in petroleum production deals with transport from the (subsea) well-bore to a topside separator. Here, a simple model in Sharma & Glemmestand (2014) has been streamlined into a dynamic model suitable for illustration of the dynamics of oil transport, as well for control studies. The advantages of using dimensionless equipment models are emphasized. The model is then used to compare two popular modeling languages: Modelica, and ModelingToolkit for Julia. Key advantages and disadvantages of these two languages are emphasized.

1 Introduction

1.1 Background

Petroleum products have been key energy carriers for more than a century. Current focus on climate[1] implies a change towards sustainable energy carriers. To succeed in this change, a transition period from the use of fossil fuel is necessary. In the transition, improved operation of petroleum production through model based optimal operation will be necessary. Petroleum production entails slow (reservoir; months) and fast (reservoir-to-separator; seconds) subsystems; a focus of research project "DigiWell"[2]. Vertical transport of petroleum from oil well to surface requires sufficient pressure to counteract gravitational and friction forces. If the oil-well heel pressure is insufficient for such transport, either (i) gas is injected in the vertical pipe to "blow" the petroleum fluids to the surface, or (ii) an electrical submersed pump [ESP] is installed in the vertical pipe to sufficiently increase the pressure. Here, we study the dynamics of transport from the reservoir formation to a surface manifold via an ESP, and further horizontal transport from the manifold to a separator.

Industrial simulation tools typically put main emphasis on the dynamics of the *reservoir* (time constant: months) and use steady state models for the reservoir-to-surface transport. This emphasis is inadequate for daily operation and control. Here, a simple dynamic reference model for oil transport from reservoir to separator is provided. The model provides an understanding of the dynamic behavior of such systems, and is suitable for industrial control design. Emphasis is put on a simple, yet stringent model development, while avoiding unit complexities.

1.2 Previous work

Sharma & Glemmestand (2014) (Sharma, 2014) provide a dynamic model of oil transport from reservoir to separator suitable for control design; this model is the focus here. Binder et al. (2015) discuss an older model; other models typically are CFD models, etc., too complex for control design. Sharma's model considers a case with 4 vertical pipes from oil reservoirs to a single manifold, with 2 horizontal pipes from the manifold to a single separator. Each vertical pipe has an ESP, plus a choke valve at the manifold entrance; the pump speeds can be manipulated individually. The horizontal pipes have booster pumps to counteract friction losses. The original ESP model includes induction motors, but the dynamics of the pump actuator is fast, and is neglected in later work. Sharma & Glemmestand (2014) provide a novel ESP model, a simple model for a booster pump, and use a valve model based on on the ANSI/ISA S75.01 standard[3].

The model with ESP in Sharma (2014) is mainly relevant for the production of heavy oil. Several papers use this model in advanced industrial control studies Krishnamoorthy et al. (2016); Santana et al. (2021).

Mixtures of liquid oil and water form an emulsion when stirred (e.g., in a multi-stage ESP); for such emulsions, the viscosity — and hence the friction — varies dramatically with water content, Justiniano & Romero (2021). Sharma & Glemmestand (2014) as-

[1] https://sdgs.un.org/goals
[2] See Acknowledgments.
[3] `http://integrated.cc/cse/ISA_750101_SPBd.pdf`

sume an unrealistic linear dependence of water fraction.

1.3 Structure of paper

Section 2 gives an overview of the transport system from oil reservoir via manifold to a separator, and key equipment models. Section 3 develops a simple mechanistic model of the system. Section 4 contrasts two modeling languages for simulation: Modelica and Julia's ModelingToolkit. Section 5 illustrates model behavior and the use of modeling/simulation tools. Finally, Section 6 provides some conclusions.

2 System description

We consider production of a mixture of water and crude oil in liquid phase.

2.1 System topology

Oil production *systems* merge several boreholes from the same or different reservoirs through vertical pipes into a manifold. Normally, more than one horizontal transport pipe are needed from the manifold to a separator for sufficient transport capacity. Water is commonly added to the manifold to reduce friction loss in the horizontal pipes. Figure 1 shows a system with n_w wells/vertical pipes and n_t transportation/horizontal pipes to the separator.
All vertical pipes are assumed connected to the same manifold pressure p_m; hence effluent *choke pressure* satisfies $p_\mathrm{c}^{\mathrm{e},j} = p_\mathrm{c}^\mathrm{e} = p_\mathrm{m}$ for all j. Likewise, all transport pipes end up in the same separator: $p_\mathrm{s}^{-,j} = p_\mathrm{s}$ for all j.

2.2 Fluid properties

The petroleum fluid properties are important. Density varies with pressure and temperature, $\rho\,(p,T)$. Neglecting temperature dependence, and assuming constant *isothermal compressibility*,

$$\rho = \rho_0 \exp\left(\beta_T\,(p - p_0)\right) \tag{1}$$

where (ρ_0, p_0) is some reference state, and β_T is the (assumed) constant isothermal compressibility.[4]
Defining water cut χ_w as $\chi_\mathrm{w} \triangleq \dot{V}_\mathrm{w}/\dot{V}$: volumetric flow rate of water divided by total flow rate of the fluid, total density ρ becomes

$$\rho = \chi_\mathrm{w}\rho_\mathrm{w} + (1-\chi_\mathrm{w})\,\rho_\mathrm{o}; \tag{2}$$

here, ρ_w and ρ_o are constant densities of pure water and crude oil, respectively.
In reality, water and crude oil have different isothermal compressibilities. Here, we simplify and assume an overall value for β_T. Using data in Appendix 1, density ρ varies ca. $10\,\mathrm{kg/m^3}$ with pressure variation in the range 25–225 bar; we thus assume constant density in pipes, but a pressure-dependent density will be assumed in the manifold.
Sharma & Glemmestad (2014) propose a simple linear mixing rule for *kinematic viscosity* ν:

$$\nu = \chi_\mathrm{w}\nu_\mathrm{w} + (1-\chi_\mathrm{w})\,\nu_\mathrm{o}. \tag{3}$$

With ν known, *dynamic* viscosity μ can be computed (if needed) as

$$\mu = \nu\rho. \tag{4}$$

This linear interpolation model is used here, even though it is not physically realistic.

2.3 Well-bore production

Total production from the reservoir (formation pressure p_f) relates volumetric petroleum fluid rate $\dot{V}_\mathrm{h}$ at the well-bore heel as $\dot{V}_\mathrm{h} \propto p_\mathrm{f} - p_\mathrm{h}$, where p_h is heel pressure and the proportionality constant is the *productivity index*, which is unit-dependent. Here, we propose a dimensionless form instead,

$$\dot{V}_\mathrm{h} = \dot{V}_\mathrm{pi}\frac{p_\mathrm{f} - p_\mathrm{h}}{p_\mathrm{pi}^\varsigma} \tag{5}$$

where $\dot{V}_\mathrm{pi}$ is the productivity *capacity* in the same unit as $\dot{V}_\mathrm{h}$ and a scaling pressure p_pi^ς which has the same unit as $p_\mathrm{f}, p_\mathrm{h}$.

2.4 Pump models

Pump models are typically given as

$$\Delta p_\mathrm{p} = \rho g h_\mathrm{p}; \tag{6}$$

pump *head* $h_\mathrm{p} = h_\mathrm{p}\left(\dot{V}, f_\mathrm{p}\right)$ with control input f_p — rotational pump frequency Hz, and volumetric flow rate $\dot{V}$.
Sharma & Glemmestad (2014) give a comprehensive model for the pump head of a *multi-stage ESP*. To ease change of units, their model is here rewritten in dimensionless form

$$\frac{h_\mathrm{p}\left(\dot{V}, f_\mathrm{p}\right)}{h_\mathrm{p}^\varsigma} = \left(\frac{f_\mathrm{p}}{f_{\mathrm{p},0}}\right)^2 + a_1\frac{f_\mathrm{p}}{f_{\mathrm{p},0}}\frac{\dot{V}}{\dot{V}^\varsigma} + a_2\left(\frac{\dot{V}}{\dot{V}^\varsigma}\right)^2 + a_3\frac{f_{\mathrm{p},0}}{f_\mathrm{p}}\left(\frac{\dot{V}}{\dot{V}^\varsigma}\right)^3. \tag{7}$$

In Eq. 7, h_p^ς is a scaling head, f_p is the pump rotational frequency in the same unit as that of the nominal rotational frequency $f_{\mathrm{p},0}$, $\dot{V}$ is the actual volumetric flow rate out of the pump, $\dot{V}^\varsigma$ a scaling flow rate, and $a_1, \ldots, a_3$ are dimensionless model parameters[5].

[4] Isothermal compressibility is the inverse of bulk modulus.

[5] Here, a_j is dimensionless, while in Sharma (2014) his parameters a_j have dimensions. This implies that the values of a_j here are different from those of a_j in Sharma (2014).

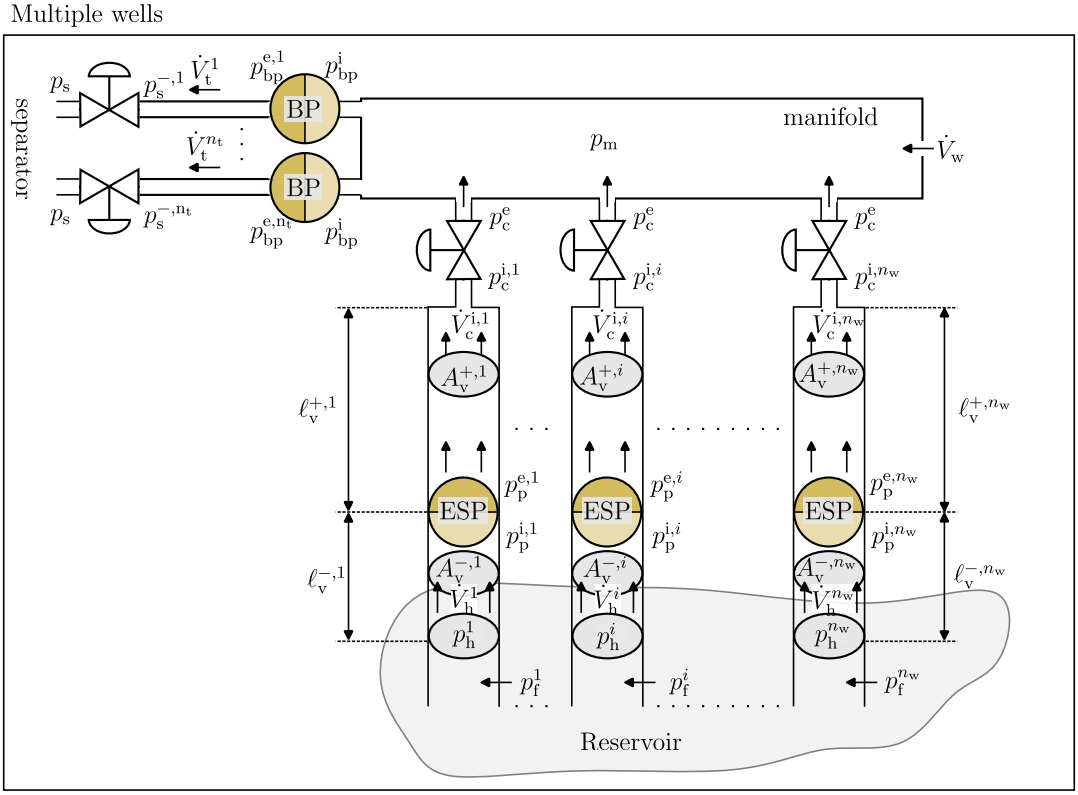


Figure 1. Multiple well system with n_w wells — possibly coming from different reservoirs, and n_t transport pipes to the separator; based on Sharma & Glemmestand (2014).

For the *booster pump* in the horizontal pipes, a simpler model is suggested in Sharma & Glemmestand (2014), here rewritten in dimensionless form as

$$\frac{\Delta p_{\mathrm{bp}}\left(f_{\mathrm{bp}}\right)}{\Delta p_{\mathrm{bp}}^{\varsigma}} = \left(\frac{f_{\mathrm{bp}}}{f_{\mathrm{bp},0}}\right)^2 \tag{8}$$

Here, $\Delta p_{\mathrm{bp}}\left(f_{\mathrm{bp}}\right)$ is the pressure increase at the given pump frequency/speed f_{bp}, in the same unit as $\Delta p_{\mathrm{bp}}^{\varsigma}$ — which is the pressure increase at the nominal pump frequency $f_{\mathrm{bp},0}$.

2.5 Valve models

Sharma & Glemmestand (2014) base their valve models on the ANSI/ISA S75.01 standard[6]. Here, we instead propose a dimensionless description with extension to a control valve as

$$\dot{m} = \dot{m}_{\mathrm{v}}^{\mathrm{c}} \cdot f\left(u_{\mathrm{v}}\right) \frac{\rho_{\mathrm{i}}}{\rho_{\mathrm{e}}} \sqrt{\frac{\left(p_{\mathrm{i}} - p_{\mathrm{e}}\right)/p^{\varsigma}}{\rho_{\mathrm{i}}/\rho^{\varsigma}}} \tag{9}$$

where $\dot{m}_{\mathrm{v}}^{\mathrm{c}}$ is the valve mass flow rate capacity, $u_{\mathrm{v}} \in [0,1]$ is the valve control signal, $f : [0,1] \to [0,1]$ is the valve characteristics, $\rho_{\mathrm{i}}, \rho_{\mathrm{e}}$ are influent and effluent densities, respectively, $p_{\mathrm{i}}, p_{\mathrm{e}}$ are influent and effluent pressures, respectively, while $\rho^{\varsigma}, p^{\varsigma}$ are scaling density and pressure, respectively.

[6] `http://integrated.cc/cse/ISA_750101_SPBd.pdf`

2.6 Friction loss

The friction drop along the pipe can be given by the Darcy-Weisbach equation,

$$\frac{\Delta p_{\mathrm{f}}}{\ell} = f_{\mathrm{D}} \frac{\rho}{2} \frac{v^2}{D} \tag{10}$$

where f_{D} is Darcy's friction factor, given by Colebrook's[7] implicit expression. One explicit approximation to Colebrook's expression is due to Swamee and Jain (Brkić, 2011),

$$\frac{1}{\sqrt{f_{\mathrm{D}}}} = -2 \cdot \log_{10}\left(\frac{5.74}{N_{\mathrm{Re}}^{0.9}} + \frac{\varepsilon/D}{3.7}\right), \tag{11}$$

where N_{Re} is the Reynolds number,

$$N_{\mathrm{Re}} = \frac{\rho v D}{\mu} = \frac{vD}{\nu}, \tag{12}$$

μ is *dynamic* viscosity, ν is *kinematic* viscosity, and ε is the "roughness height" of the pipe internals. Linear velocity v is related to volumetric flow rate $\dot{V}$ by

$$\dot{V} = vA \tag{13}$$

where A is the cross-sectional area of the pipe.

[7] The Colebrook equation, or sometimes known as the Colebrook-White equation.

2.7 Why dimensionless models?

As an example, consider the ESP model in Eq. 7. In the original formulation in Sharma (2014), the volumetric flow-rate is hard-coded to use a given unit for the flow rate, e.g., m^3/day. If the dynamic model requires the flow rate in other units for dimensional consistency, it may take considerable work to re-compute the polynomial coefficients to achieve this. In summary: use of dimensionless models simplifies the process of changing units, and reduces the chance of introducing errors.

3 Dynamic model

3.1 Balance laws

The model is based on the total mass balance (manifold) and the linear momentum balance (pipes). The total mass balance is expressed as

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \dot{m}_\mathrm{i} - \dot{m}_\mathrm{e} \tag{14}$$

where m is accumulated mass in the system, t is time, $\dot{m}$ is mass flow rate, and indices i, e denote influent and effluent, respectively.
The linear momentum balance is

$$\frac{\mathrm{d}\mathfrak{m}}{\mathrm{d}t} = \dot{\mathfrak{m}}_\mathrm{i} - \dot{\mathfrak{m}}_\mathrm{e} + F, \tag{15}$$

where $\mathfrak{m}$ is linear momentum given as $\mathfrak{m} = mv$ with linear velocity v, $\dot{\mathfrak{m}}$ is momentum flow rate given as $\dot{\mathfrak{m}} = \dot{m}v$, and F is total force. With constant fluid density, $\dot{\mathfrak{m}}_\mathrm{i} = \dot{\mathfrak{m}}_\mathrm{e}$, and the momentum balance reduces to Newton's law, $\frac{\mathrm{d}\mathfrak{m}}{\mathrm{d}t} = F$.

3.2 Vertical pipes with ESP

We assume constant density in the pipes, hence Eq. 15 reduces to Newton's law. Momentum is given as $\mathfrak{m} = mv$ with $m = \rho\dot{V}$, and v related to $\dot{V}$ by Eq. 13. The total force is $F = F_\mathrm{p} + F_\mathrm{b} - F_\mathrm{f} - F_\mathrm{g}$, with

- Pressure forces at inlet and outlet of the pipe,

$$F_\mathrm{p} = p_\mathrm{h}A - p_\mathrm{c}^\mathrm{i}A \tag{16}$$

- Possible pressure boost due to a pump,

$$F_\mathrm{b} = \Delta p_\mathrm{p} A, \tag{17}$$

  with Δp_p given by Eqs. 6, 7,

- Friction loss,

$$F_\mathrm{f} = \Delta p_\mathrm{f} A, \tag{18}$$

  with Δp_f given by Eqs. 10, 11, 12, 13,

- Flow against gravity, with a lift height h,

$$F_\mathrm{g} = \Delta p_\mathrm{g} A, \tag{19}$$

  with

$$\Delta p_\mathrm{g} = \rho_\mathrm{v} g h.$$

In addition, we need information about how flow rate $\dot{V}$ relates to the bottom hole pressure via the productivity capacity, Eq. 5.
The most structured formulation would be to pose the momentum balance (here: Newton's law) as the differential equation, and add all necessary algebraic equations. However, the OpenModelica DAE solver struggles with such a formulation: the valve equation Eq. 9 is implicit in pressure difference; in the iteration to find $\Delta p_\mathrm{v} = p_\mathrm{i} - p_\mathrm{e}$, if Δp_v becomes negative, the square root gives a complex number, and the simulation crashes. Instead, we have changed the differential variable to $\dot{V}$; then the valve equation can be inverted and expressed as $\Delta p_\mathrm{v} \propto \dot{V}^2$.
The following formulation is used in OpenModelica and ModelingToolkit:

$$\frac{\mathrm{d}\dot{V}_\mathrm{v}}{\mathrm{d}t} = \frac{p_\mathrm{h} - p_\mathrm{i}^\mathrm{c} + \Delta p_\mathrm{p} - \Delta p_\mathrm{f} - \Delta p_\mathrm{g}}{\rho_\mathrm{v}\ell/A_\mathrm{v}} \tag{20}$$

$$\rho_\beta^0 = \chi_\mathrm{w}\rho_\mathrm{w} + (1-\chi_\mathrm{w})\,\rho_\mathrm{o} \tag{21}$$

$$\nu = \chi_\mathrm{w}\nu_\mathrm{w} + (1-\chi_\mathrm{w})\,\nu_\mathrm{w} \tag{22}$$

$$\mu = \rho_\beta^0 \nu \tag{23}$$

$$\rho_\mathrm{v} = \rho_\beta^0 \exp\left(\beta_T\left(p_\mathrm{c}^\mathrm{i} - p_\beta^0\right)\right) \tag{24}$$

$$p_\mathrm{h} = p_\mathrm{f} - p_\mathrm{pi}^\varsigma \frac{\dot{V}_\mathrm{v}}{\dot{V}_\mathrm{pi}} \tag{25}$$

$$\dot{m}_\mathrm{v} = \rho_\mathrm{v}\dot{V}_\mathrm{v} \tag{26}$$

$$p_\mathrm{i}^\mathrm{c} = p_\mathrm{m} + p_\mathrm{v}^\varsigma \frac{\rho_\mathrm{v}}{\rho_\mathrm{v}^\varsigma}\left(\frac{\dot{m}_\mathrm{v}}{\dot{m}_\mathrm{v}^\mathrm{c}}\right)^2 \tag{27}$$

$$h_\mathrm{p} = h_\mathrm{p}^\varsigma\left(\left(\frac{f_\mathrm{p}}{f_{\mathrm{p},0}}\right)^2 + a_1\frac{f_\mathrm{p}}{f_{\mathrm{p},0}}\frac{\dot{V}}{\dot{V}^\varsigma} + a_2\left(\frac{\dot{V}}{\dot{V}^\varsigma}\right)^2 + a_3\frac{f_{\mathrm{p},0}}{f_\mathrm{p}}\left(\frac{\dot{V}}{\dot{V}^\varsigma}\right)^3\right) \tag{28}$$

$$\Delta p_\mathrm{p} = \rho_\mathrm{v} g h_\mathrm{p} \tag{29}$$

$$v_\mathrm{v} = \frac{\dot{V}_\mathrm{v}}{A} \tag{30}$$

$$N_\mathrm{Re} = \frac{\rho_\mathrm{v} v_\mathrm{v} d_\mathrm{v}}{\nu_\mathrm{v}} \tag{31}$$

$$f_\mathrm{D}^\mathrm{v} = \frac{1}{4\left(\log_{10}\left(\frac{5.74}{N_\mathrm{Re}^{0.9}} + \frac{\varepsilon_\mathrm{v}/d_\mathrm{v}}{3.7}\right)\right)^2} \tag{32}$$

$$\Delta p_{\mathrm{f}} = \ell \cdot f_{\mathrm{D}}^{\mathrm{v}} \frac{\rho_{\mathrm{v}}}{2} \frac{v_{\mathrm{v}}^2}{d_{\mathrm{v}}} \tag{33}$$

$$\Delta p_{\mathrm{g}} = \rho_{\mathrm{v}} g h. \tag{34}$$

If we only consider the model of a single vertical pipe, we need to specify (i) initial state (i.e., $\dot{V}_{\mathrm{v}}$), (ii) all "input" variables, i.e., p_{f}, f_{p}, p_{m}, and possibly water cut χ_{w}, and (iii) all parameters, i.e., ρ_{w}, ρ_{o}, ν_{w}, ν_{o}, p_β^0, ℓ, A, $p_{\mathrm{pi}}^\varsigma$, $\dot{V}_{\mathrm{pi}}$, p_{v}^ς, $\rho_{\mathrm{v}}^\varsigma$, $\dot{m}_{\mathrm{v}}^{\mathrm{c}}$, h_{p}^ς, $f_{\mathrm{p},0}$, $\dot{V}^\varsigma$, a_1,a_2, a_3, g, d_{v}, ν_{v}, ε_{v}, h.

3.3 Manifold

We assume a perfectly mixed manifold. Assuming constant manifold volume V_{m}, and adding water at flow rate $\dot{V}_{\mathrm{w}}$ to dilute the fluid to manifold water cut $\chi_{\mathrm{w}}^{\mathrm{m}}$, thus reducing friction loss in the pipe towards separator, $\dot{V}_{\mathrm{w}}$ must be approximately

$$\dot{V}_{\mathrm{w}} = \frac{\chi_{\mathrm{w}}^{\mathrm{m}} - \chi_{\mathrm{w}}}{1 - \chi_{\mathrm{w}}^{\mathrm{m}}} \dot{V}_{\mathrm{v}}. \tag{35}$$

Total mass balance for the manifold can then be expressed as

$$\frac{\mathrm{d}p_{\mathrm{m}}}{\mathrm{d}t} = \frac{1}{\rho_{\mathrm{m}} V_{\mathrm{m}} \beta_T} \left(\rho_{\mathrm{v}} \dot{V}_{\mathrm{v}} + \rho_{\mathrm{w}} \dot{V}_{\mathrm{w}} - \rho_{\mathrm{m}} \dot{V}_{\mathrm{t}} \right) \tag{36}$$

$$\rho_\beta^0 = \chi_{\mathrm{w}}^{\mathrm{m}} \rho_{\mathrm{w}} + (1 - \chi_{\mathrm{w}}^{\mathrm{m}}) \rho_{\mathrm{o}} \tag{37}$$

$$\rho_{\mathrm{m}} = \rho_\beta^0 \exp\left(\beta_T \left(p_{\mathrm{m}} - p_\beta^0 \right) \right) \tag{38}$$

$$\dot{V}_{\mathrm{w}} = \frac{\chi_{\mathrm{w}}^{\mathrm{m}} - \chi_{\mathrm{w}}}{1 - \chi_{\mathrm{w}}^{\mathrm{m}}} \dot{V}_{\mathrm{c}}^{\mathrm{i}} \tag{39}$$

In practice, a control system must be used to manipulate $\dot{V}_{\mathrm{w}}$ instead of using Eq. 35.

For the manifold model, we must know (i) the initial manifold pressure, (ii) the vertical inflow $\dot{V}_{\mathrm{v}}$ and the horizontal transport flow $\dot{V}_{\mathrm{t}}$ from manifold to separator, as well as manifold water cut $\chi_{\mathrm{w}}^{\mathrm{m}}$, and (iii) parameters.

3.4 Transport pipe with booster pump

The model of the horizontal pipe from manifold to separator is almost identical to the vertical pipe from reservoir to manifold. The essential differences are (i) no gravity pressure drop, (ii) simpler booster pump model, (iii) neglecting pressure drop from pipe into separator, (iv) no need for a production capacity model. The complete model is

$$\frac{\mathrm{d}\dot{V}_{\mathrm{t}}}{\mathrm{d}t} = \frac{p_{\mathrm{m}} - p_{\mathrm{s}} + \Delta p_{\mathrm{bp}} - \Delta p_{\mathrm{f}}^{\mathrm{t}}}{\rho_{\mathrm{t}} \ell_{\mathrm{t}} / A_{\mathrm{t}}} \tag{40}$$

$$\rho_\beta^{0,\mathrm{t}} = \chi_{\mathrm{w}}^{\mathrm{m}} \rho_{\mathrm{w}} + (1 - \chi_{\mathrm{w}}^{\mathrm{m}}) \rho_{\mathrm{o}} \tag{41}$$

$$\nu_{\mathrm{t}} = \chi_{\mathrm{w}}^{\mathrm{m}} \nu_{\mathrm{w}} + (1 - \chi_{\mathrm{w}}^{\mathrm{m}}) \nu_{\mathrm{w}} \tag{42}$$

$$\mu_{\mathrm{t}} = \rho_\beta^{0,\mathrm{t}} \nu_{\mathrm{t}} \tag{43}$$

$$\rho_{\mathrm{t}} = \rho_\beta^0 \exp\left(\beta_T \left(p_{\mathrm{m}} - p_\beta^0 \right) \right) \tag{44}$$

$$\Delta p_{\mathrm{bp}} = \Delta p_{\mathrm{bp}}^\varsigma \left(\frac{f_{\mathrm{bp}}}{f_{\mathrm{bp},0}} \right)^2 \tag{45}$$

$$v_{\mathrm{t}} = \frac{\dot{V}_{\mathrm{t}}}{A_{\mathrm{t}}} \tag{46}$$

$$N_{\mathrm{Re,t}} = \frac{\rho_{\mathrm{t}} v_{\mathrm{t}} d_{\mathrm{t}}}{\nu_{\mathrm{t}}} \tag{47}$$

$$f_{\mathrm{D}}^{\mathrm{t}} = \frac{1}{4 \left(\log_{10} \left(\frac{5.74}{N_{\mathrm{Re,t}}^{0.9}} + \frac{\varepsilon_{\mathrm{t}}/d_{\mathrm{t}}}{3.7} \right) \right)^2} \tag{48}$$

$$\Delta p_{\mathrm{f}}^{\mathrm{t}} = \ell_{\mathrm{t}} \cdot f_{\mathrm{D}}^{\mathrm{t}} \frac{\rho_{\mathrm{t}}}{2} \frac{v_{\mathrm{t}}^2}{d_{\mathrm{t}}}. \tag{49}$$

Again, we need to know the initial condition of the differential variable ($\dot{V}_{\mathrm{t}}$), the inputs ($\chi_{\mathrm{w}}^{\mathrm{m}}$, f_{bp}, p_{m}, p_{s}), and the parameters.

3.5 Combined model

For illustration, we use two vertical pipes, one manifold, and one horizontal transport pipe from manifold to separator. Both Modelica and Julia's ModelingToolkit have support for building classes/reusable models. Because of the similarity between the models for vertical and horizontal pipes, it would be possible to collect these in the same class/constructor and just differentiate between them with a function argument. The manifold model should be a separate class, though.

With re-usability of such classes/constructors, modeling of the combined system simply consists of (i) instantiating one model per unit (2 vertical pipes, one horizontal transport pipe, and the manifold), and (ii) connecting the various instances. Specifically, the vertical pipes should see the same manifold pressure p_{m}, the vertical transport pipe should have the same inlet pressure as the manifold pressure p_{m}, the influent volumetric flows to the manifold should be the sum of the flows from the vertical pipes and the viscosity diluting water feed $\dot{V}_{\mathrm{w}}$ now being

$$\dot{V}_{\mathrm{w}} = \frac{\sum_{i=1}^{2} \left(\chi_{\mathrm{w}}^{\mathrm{m}} - \chi_{\mathrm{w}}^{i} \right) \dot{V}_{\mathrm{v}}^{i}}{1 - \chi_{\mathrm{w}}^{\mathrm{m}}}; \tag{50}$$

the effluent volumetric flow from the manifold is still $\dot{V}_{\mathrm{t}}$.

For a proper re-usable implementation, connections should be done using *connectors* (supported by both Modelica and ModelingToolkit).

4 Simulation tools

The combined model has been solved using the free languages/tools OpenModelica (Fritzson, 2015; Fritzson et al., 2018) and ModelingToolkit (Ma et al., 2021) for Julia; the results are identical plus/minus variations due to solver accuracies. Results presented in Section 5 use the ModelingToolkit/Julia implementation due to better support in Julia for plotting and random variables.

Modelica is a mature language dating back to the 1990s; ModelingToolkit is some 2–3 years young and is still in some flux. ModelingToolkit is evolving rapidly, is more general than Modelica, and is also integrated in the larger Eco-system of Julia. Currently, ModelingToolkit does not support a graphical flow-sheeting tool, and it is unclear whether ModelingToolkit allows for as large models as OpenModelica. Both tools have extensive support for building libraries.

The default solver in OpenModelica is excellent, although here it struggled with the DAE formulation with momentum as differential variable. ModelingToolkit can use solvers from the large, high quality DifferentialEquations.jl package (Rackauckas & Nie, 2017). With ModelingToolkit, choice of solver, accuracies, etc., currently requires more thought compared to OpenModelica. The solutions from ModelingToolkit include interpolation functions, which yields smooth solutions with fewer data points.

OpenModelica normally works well when providing initial conditions for differential variables only, while with ModelingToolkit it is necessary to also specify initial values for algebraic variables.

OpenModelica's support for linearization and plotting can be accessed from Julia via the OMJulia API (B. Lie et al., 2019). ModelingToolkit is integrated in the Julia Eco-system, with support for linearization, plotting, control systems analysis, random variables, etc., and has overall more possibilities that OpenModelica if further analysis is required.

Other commonly used languages for scientific computing are MATLAB (commercial) and Python (free). Compared to both of these languages, Julia (free) has a more extensive set of differential equation solvers. Neither MATLAB nor Python offer equation based modeling languages with library/re-use support such as Modelica or ModelingToolkit; MathWorks do offer Simscape[8] (commercial) for such use, though.

[8] https://se.mathworks.com/products/simscape.html

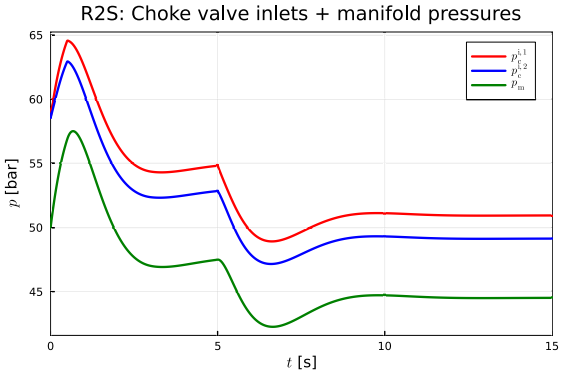


Figure 2. Pressures in front of choke valve into manifold for vertical pipes (red, blue) and manifold pressure (green).

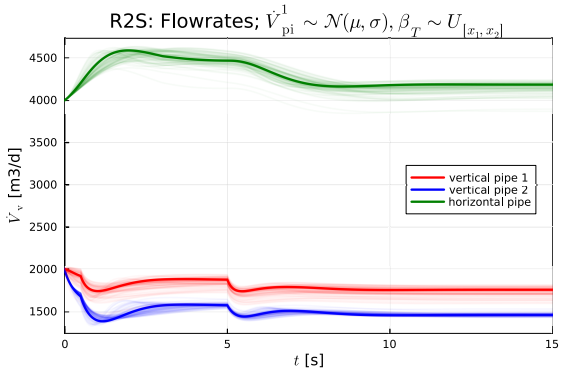


Figure 3. Vertical flow rates (red, blue) from bore-well into manifold, and horizontal flow rate (green) from manifold to separator, with uncertainty productivity capacity and isothermal compressibility.

5 Results

Parameters, initial conditions, and system inputs are given in Appendix A. For vertical pipe #2, scaling pump head $h^{\varsigma}_{\mathrm{p}}$ is set to 80% of the value suggested in the appendix. Figure 2 shows the pressures in front of the choke valves for the vertical pipes, as well as the manifold pressure. The resulting time constants and overall behavior in Fig. 2 are similar to those in Sharma (2014).

Figure 3 shows vertical flow rates from reservoir to manifold in the two pipes, as well as the flow from manifold to separator (thick, solid lines), and the effect of uncertain productivity indices in Well 1, $\dot{V}^1_{\mathrm{pi}} \sim \mathcal{N}\left(7\cdot 10^{-4}, 10^{-4}\right)$, and uncertain isothermal compressibility in the petroleum fluid, $\beta_T \sim U_{\left[0.3/1.5\cdot 10^9, 3/1.5\cdot 10^9\right)}$.

ModelingToolkit has support for efficient Monte Carlos studies; this is comparatively more complicated using Modelica + OMJulia.

Both Modelica+OMJulia and ModelingToolkit have similar possibilities to linearize the models, and Con-

trolSystems.jl for Julia has similar capabilities as MATLAB's Control Toolbox for plotting and analysis/design.

6 Conclusions

This paper presents a simple model of production of liquid petroleum (oil+water) from reservoir to separator. The model is essentially a reworking of the model in Sharma & Glemmestand (2014). Modifications include: (i) a stricter utilization of the constant density assumption in pipes[9] leading to a more realistic behavior at choke valves, (ii) rephrasing of algebraic equipment models into dimensionless form, which greatly simplifies unit conversion, (iii) streamlining of the model presentation to ease the implementation of the model; in the original formulation, some information is missing, and information is spread through a long paper, (iv) scaling down the model from 4 vertical pipes/2 horizontal pipes to 2 vertical pipes/1 horizontal pipe.

The model is implemented in OpenModelica and in Julia with ModelingToolkit. These tools have similar capabilities, although Modelica is more mature, has perhaps better default solver, and can handle larger systems at the moment. However, ModelingToolkit is embedded in the larger Eco-system of Julia, with superior capabilities for plotting, uncertainty analysis[10], simpler linearization, control analysis and design, etc. Combining OpenModelica with OMJulia, some of the features of the Julia Eco-system can be utilized (plotting, linearization, etc.). However, with ModelingToolkit, other tools in Julia have access to the symbolic form of the model, and can symbolically compute Jacobians, etc. Both of the free tools OpenModelica and ModelingToolkit are equation based modeling languages with solid support for model libraries and re-use of code.

The presented model was developed for short-term industrial oil production control design, Sharma (2014). More comprehensive models typically include a reservoir model (time constant: months+) suitable for long-term simulation studies (K.-A. Lie, 2019), with a steady state network solver for the transport from reservoir to separator (time constant: seconds+), thereby avoiding stiffness issues. These steady state models are not really suitable for control design for daily operation, while the model presented here has been used to assess industrial control policies.

A number of possible extensions for the system include (a) more realistic properties (density, viscosity), (b) allowing for distributed density along pipes[11], (c) adding a more realistic system for water dilution in the manifold, (d) inclusion of valves in manifold–separator pipes, (e) integration with reservoir models, (f) use for control design, (g) use for optimization, etc. Such extensions will give more insight into the industrial usefulness of the model.

[9] The original model includes differential equation for the pipe mass balance, although the mass is assumed constant.

[10] Modelica lacks proper support for random numbers.

[11] ModelingToolkit for Julia has support for automatic discretization of PDEs in the works.

Table 1. Parameters: petroleum liquid.

Parameter
$\beta_T = \frac{1}{1.5\cdot 10^9} \approx 6.67\cdot 10^{-10}\,\mathrm{Pa}^{-1}$
$p_0 = 1\,\mathrm{bar}$
$\rho_\mathrm{o} = 900\,\mathrm{kg/m^3}$
$\rho_\mathrm{w} = 1000\,\mathrm{kg/m^3}$
$\chi_\mathrm{w} = 0.35$
$\rho_0 = \chi_\mathrm{w}\rho_\mathrm{w} + (1-\chi_\mathrm{w})\,\rho_\mathrm{o}$
$\chi_\mathrm{w}^\mathrm{m} = 0.5$
$\rho_0^\mathrm{m} = \chi_\mathrm{w}^\mathrm{m}\rho_\mathrm{w} + (1-\chi_\mathrm{w}^\mathrm{m})\,\rho_\mathrm{o}$
$\nu_\mathrm{o} = 100\,\mathrm{cSt} = 100\cdot 10^{-6}\,\mathrm{m^2/s}$
$\nu_\mathrm{w} = 1\,\mathrm{cSt} = 10^{-6}\,\mathrm{m^2/s}$

—Acknowledgments

The economic support from The Research Council of Norway and Equinor ASA through Research Council project "308817 —Digital wells for optimal production and drainage" (DigiWell) is gratefully acknowledged.

A Parameters and Operating Conditions

Parameters for petroleum fluid, nominal vertical pipes, and nominal manifold+horizontal pipe are given in Tables 1–3. Initial states are given in Table 4, while input functions are given in Table 5.

References

Binder, B. J. T., Pavlov, A., & Johansen, T. A. (2015). Estimation of flow rate and viscosity in a well with an electric submersible pump using moving horizon estimation. *IFAC-PapersOnLine*, *48-6*, 140–146.

Brkić, D. (2011, April). Review of explicit approximations to the Colebrook relation for flow friction. *Journal of Petroleum Science and Engineering*, *77*(1), 34-48. (http://dx.doi.org/10.1016/j.petrol.2011.02.006)

Fritzson, P. (2015). *Principles of Object-Oriented Modeling and Simulation with Modelica 3.3: A Cyber-Physical Approach* (Second ed.). Piscataway, NJ: Wiley-IEEE Press.

Fritzson, P., Pop, A., Asghar, A., Bachmann, B., Braun, W., Braun, R., … Östlund, P. (2018, October). The OpenModelica Integrated Modeling, Simulation and Optimization Environment. In

Table 2. Parameters: vertical pipe.

Parameter
$\ell^- = 100\,\mathrm{m}$
$\ell^+ = 2000\,\mathrm{m}$
$d = 0.1569\,\mathrm{m}$
$\varepsilon = 0.0018\,\mathrm{inch} = 45.7\,\mu\mathrm{m} = 45.7 \cdot 10^{-6}\,\mathrm{m}$
$h_\mathrm{p}^\varsigma = 1210.6\,\mathrm{m}$
$f_{\mathrm{p},0} = 60\,\mathrm{Hz}$
$\dot{V}^\varsigma = 1\,\mathrm{m}^3/\mathrm{s}$
$a_1 = -37.57$
$a_2 = 2.864 \cdot 10^3$
$a_3 = -8.668 \cdot 10^4$
$\dot{m}_\mathrm{v}^\mathrm{c} = 25.9 \cdot 10^3\,\mathrm{kg/h}$
$f(u_\mathrm{v}) = \begin{cases} 0, & u_\mathrm{v} \le 0.05 \\ \frac{11.1u_\mathrm{v} - 0.556}{30}, & 0.05 < u_\mathrm{v} \le 0.5 \\ \frac{50u_\mathrm{v} - 20}{30}, & 0.5 < u_\mathrm{v} \le 1 \end{cases}$
$p^\varsigma = 1\,\mathrm{bar}$
$\rho^\varsigma = 1000\,\mathrm{kg/m}^3$
$\dot{V}_\mathrm{pi} = 7 \cdot 10^{-4}\,\mathrm{m}^3/\mathrm{s}$

Table 3. Parameters: manifold+horizontal pipe.

Parameter
$\ell_\mathrm{m} = 500\,\mathrm{m}$
$d_\mathrm{m} = 0.1569\,\mathrm{m}$
$\ell_\mathrm{t} = 4000\,\mathrm{m}$
$d_\mathrm{t} = 0.1569\,\mathrm{m}$
$\varepsilon = 0.0018\,\mathrm{inch} = 45.7\,\mu\mathrm{m} = 45.7 \cdot 10^{-6}\,\mathrm{m}$
$\Delta p_\mathrm{bp}^\varsigma = 10\,\mathrm{bar}$
$f_{\mathrm{bp},0} = 60\,\mathrm{Hz}$

Table 4. Nominal initial states.

Variable
$\dot{V}_\mathrm{v}(t=0) = 2000\,\mathrm{m}^3/\mathrm{d} \approx 0.02315\,\mathrm{m}^3/\mathrm{s}$
$p_\mathrm{m}(t=0) = 50\,\mathrm{bar} = 50 \cdot 10^5\,\mathrm{Pa}$
$\dot{V}_\mathrm{t}(t=0) = 2000\,\mathrm{m}^3/\mathrm{d} \approx 0.02315\,\mathrm{m}^3/\mathrm{s}$

Table 5. Nominal inputs.

Variable
$p_\mathrm{f}(t) = \begin{cases} 220\,\mathrm{bar}, & t < 0.5\,\mathrm{s} \\ 0.95 \cdot 220\,\mathrm{bar}, & t \ge 0.5\,\mathrm{s} \end{cases}$
$p_\mathrm{s}(t) = \begin{cases} 30\,\mathrm{bar}, & t < 3\,\mathrm{s} \\ 0.97 \cdot 30\,\mathrm{bar}, & t \ge 3\,\mathrm{s} \end{cases}$
$f_\mathrm{p}(t) = \begin{cases} 60\,\mathrm{Hz}, & t < 5\,\mathrm{s} \\ 0.95 \cdot 60\,\mathrm{Hz}, & t \ge 5\,\mathrm{s} \end{cases}$
$u_\mathrm{v}(t) = 1.0$
$f_\mathrm{bp} = 60\,\mathrm{Hz}$

Proceedings of the 1st american modelica conference. Cambridge, MA, USA: LIU Electronic Press, www.ep.liu.se. (October, 8-10, 2018)

Justiniano, M., & Romero, O. J. (2021). Inversion Point of Emulsions as a Mechanism of Head Loss Reduction in Onshore Pipeline Heavy Oil Flow. *Brazilian Journal of Petroleum and Gas*, *15*(1–2), 13–24.

Krishnamoorthy, D., Bergheim, E. M., Pavlov, A., Fredriksen, M., & Fjalestad, K. (2016). Modelling and robustness analysis of model predictive control for electrical submersible pump lifted heavy oil wells. *IFAC-PapersOnLine*, *49-7*, 544–549.

Lie, B., Palanisamy, A., Mengist, A., Buffoni, L., Sjölund, M., Asghar, A., … Fritzson, P. (2019, February). OMJulia: An OpenModelica API for Julia-Modelica Interaction. In *Proceedings of the 13th international modelica conference* (pp. 699–708). (Regensburg, Germany, March 4–6, 2019) doi: 10.3384/ecp19157

Lie, K.-A. (2019). *An introduction to reservoir simulation using matlab/gnu octave: User guide for the matlab reservoir simulation toolbox (mrst)*. Cambridge University Press. doi: 10.1017/9781108591416

Ma, Y., Gowda, S., Anantharaman, R., Laughman, C., Shah, V., & Rackauckas, C. (2021). ModelingToolkit: A Composable Graph Transformation System For Equation-Based Modeling. *arXiv*. doi: 10.48550/arXiv.2103.05244

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

Santana, B. A., Fontes, R. M., Schnitman, L., & Martins, M. A. F. (2021). An adaptive infinite horizon model predictive control strategy applied to an esp-lifted oil well system. *IFAC PapersOnLine*, *54-3*, 176-181.

Sharma, R. (2014). *Optimal Operation of Gas Lifted and ESP Lifted Oil Fields: An Approach Based on Modeling, Simulation, Optimization and Control* (Unpublished doctoral dissertation). University of South-Eastern Norway, Kjølnes Ring 56, N-3918 Porsgrunn, Norway.

Sharma, R., & Glemmestand, B. (2014). Modeling and simulation of an electric submersible pump lifted oil field. *International Journal of Petroleum Science and Technology*, *8*(1), 39–68.

A Python-based code for modeling the thermodynamics of the vapor compression cycle applied to residential heat pumps

Rebecca Allen[a,*], Eirik Svortevik[a], Henrik Bergersen[a]

[a]*Oslo Metropolitan University, Norway*
[*]rebeccaa@oslomet.no

Abstract

Heat pumps are an attractive heating system in residential buildings. They operate based on the vapor compression cycle used in refrigeration systems. Design questions surrounding heat pumps can be investigated and answered using modelling tools that incorporate the necessary thermodynamics, fluid mechanics, and machinery component efficiency. Several modelling tools are available, however there is a need for more open-source, script-based programs that are competitive to those already available. This work presents a Python-based code for modeling the thermodynamics of the vapor compression cycle (VCC) in typical heat pumps. The main contribution of this work is an openly available online code, complete with a few examples to show its functionality, that provides the basic thermodynamic model of a heat pump for researchers or development engineers to use, modify, and extend. Its current features include choice of refrigerant, heat exchanger size and characteristics, compressor, and other design parameters such as heating load, and fluid temperatures in and out of the heat exchangers. Simulation outputs include the P-h and T-s diagrams and coefficient of performance (COP). The code is flexible and suggestions for future code development are given.

1. Introduction

Residential heat pumps are an attractive alternative to electrical space heating because more units of heat energy can be transferred while consuming the same amount of electrical power. Indeed, heat pump technology is not new: decades worth of research efforts have been done by, for example, the International Energy Agency (IEA) Heat Pumping Technologies (HPT). Since the heat pump cycle and refrigeration cycle are essentially the same cycle (just with different objective: one for heating, the other for cooling), it is more concise to use the term vapor-compression cycle, or VCC. The VCC is comprised of 4 main thermodynamic processes, and the central part is to transfer heat energy from a low-temperature reservoir to a high-temperature reservoir. An excellent sketch of the components is presented in Figure 2 of Jensen et al. (2018). Heat is "pumped" by first compressing a working fluid (also called refrigerant) in its vapor phase which increases its pressure and temperature, and then exchanging heat energy to another fluid in a secondary loop via both temperature drop and phase change through a condenser. The refrigerant, in its liquid phase, is then throttled, which means it passes through a throttle valve (also known as expansion valve), reducing its pressure and temperature until it becomes a saturated mixture. The mixture then passes through an evaporator where it absorbs heat energy from another secondary fluid loop and comes out in its initial vapor phase state.

The thermodynamics are well-understood in the VCC, however there is a need for flexible tools that can model these thermodynamic processes with the purpose of answering research questions such as optimal cycle "position" and "lift" given certain operating conditions. Other programs and codes have been developed and are available (Aulicino and Bakrania, 2022; Bell *et al*., 2014; Vering *et al*., 2022). However there is a need for more models that are open-source, script-based (for example, JavaScript or Python) with a framework that allows for application to one's own engineering design problem in addition to extendibility. The programming tools cannot be like a "black-box". And good documentation of code, and guidance on how to extend or modify the code ought to be available.

The objective of this work is to start an open-source repository for VCC modelling using Python scripting language. There is a great deal of potential when coding in Python, given its wide online community, wealth of libraries and modules or packages such as optimization and machine learning ones, and the fact that some engineering

companies choose to develop heat pump dimensioning tools using Python.

The rest of this paper is organized as follows: the methodology section presents the typical modelling equations used to simulate the refrigeration cycle and a description of how the model is implemented in Python, the results section presents three examples to demonstrate the use of the code and gives suggestions for how it can be extended, and the summary section offers ways this code could be developed further and applied to research on heat pump modelling.

2. Methodology

This work includes three examples that illustrate how a user could run the code with different objectives. Example 1 is based on the assumption that condenser temperature (Tcond), evaporation temperature (Tevap), degree of superheating (SH), degree of subcooling (SC), refrigerant (fluid), and compressor efficiency (n_{comp}) are known input values. From these 6 parameters, the state points in the vapor compression cycle is solved directly by the VCC-calculator. Example 2 is similar except that compressor efficiency is not assumed to be known but rather more realistically depends on the pressure ratio between the 2 state points partially defined by Tcond and Tevap. Example 3 is different from the first two examples in that it shows how the VCC-calculator can be used in a conditional while-loop that repeatedly gets called until the energy transfer from a source fluid, energy transfer across the heat pump, and energy transfer to a sink fluid all converge in the sense that the cycle yields a desired heat transfer $\dot{Q}_{sink}$ while satisfying all thermodynamic modelling equations within some set tolerance.

Programming of the VCC-calculator is based on a set of assumptions which are typically used in literature (for example, Ouadha et al., 2008, Camdali 2015, Madessa et al., 2017, Jensen et al., 2018, Wang et al., 2022) when modelling the thermodynamics of a refrigeration cycle, either in heating or cooling mode. These assumptions are:

- A steady-state, closed system with only energy (heat, work) transfer
- No heat losses between system and surroundings
- Isobaric heat exchange in the condenser and evaporator
- Perfect heat transfer between the refrigerant in the primary loop and the fluids in the secondary loops
- Isenthalpic process through the expansion (or throttle) valve
- An irreversible process through the compressor, quantified in terms of an isentropic efficiency value, n_s

The steady-state assumption implies that the mass flow rate of the refrigerant, $\dot{m}$, is constant at any point in the system. This is why the subscripts for this variables is dropped since it does not differ across the cycle's state points. No heat losses between the system and the surroundings imply that:

$$\dot{Q}_{cond} = \dot{W}_{comp} + \dot{Q}_{evap} \quad (1)$$

according to the first law of thermodynamics. The isobaric assumption means that heat transfer across the heat exchangers is given by:

$$\dot{Q}_{cond} = \dot{m}\Delta h_{cond} = \dot{m}(h_2 - h_3) \quad (2)$$

and

$$\dot{Q}_{evap} = \dot{m}\Delta h_{evap} = \dot{m}(h_1 - h_4) \quad (3)$$

Perfect heat transfer between the refrigerant and the secondary fluids (air, water, or brine) implies that $\dot{Q}_{evap} = \dot{Q}_{source}$ and $\dot{Q}_{cond} = \dot{Q}_{source}$, where the source and sink fluids release and absorb the heat according to:

$$\dot{Q}_{source} = \dot{m}_{source} c_p \Delta T \quad (4)$$

and

$$\dot{Q}_{sink} = \dot{m}_{sink} c_p \Delta T \quad (5)$$

Since the process through the expansion (or throttle) valve is assumed to be isenthalpic, the enthalpies at the state points before and after the expansion valve are equal, i.e.,

$$h_4 = h_3 \quad (1)$$

The degree of irreversibility of the process through the compressor is directly related to the compressor efficiency. According to the second-law of thermodynamics, an irreversible process is one in which the entropy difference between the initial and final state is greater and zero. That is, $\Delta s_{comp} > 0$. The specific enthalpy of the fluid that comes out from the compressor (labelled as state point 2) is thus calculated by:

$$h_2 = h_1 - \frac{h_{2s} - h_1}{n_s} \quad (6)$$

which comes from the definition of isentropic efficiency, n_s, and where h_{2s} is the specific

enthalpy at state point 2 as if the process was reversible (i.e., $s_1 = s_2$). In this work, the mechanical and electrical efficiencies of the compressor are both assumed to be 1, thus the overall work of the compressor is $\dot{W}_{comp} = \dot{m}\Delta h_{comp}$. However, if one wanted to include mechanical and electrical efficiencies, n_m and n_e, respectively, they can implement $\dot{W}_{comp} = \dot{m}\Delta h_{comp}/(n_m n_e)$ in the code.

Figure 1 shows the *myVCCmodel* definition. This is the central part of this work's VCC-repository. That is, the main state points of the VCC is obtained by this definition based on values for Tcond, Tevap, SH, SC, compressor efficiency, and fluid, which are passed in as known input parameters. This definition is based on thermodynamic principals, assumptions about the VCC already listed, and state point properties obtained using the CoolProp thermophysical property library (Bell et al., 2014). Since the CoolProp library is used in this work's code, users must first download and install CoolProp before running any example that uses the *myVCCmodel* definition. Download and installation instructions can be found online at coolprop.org.

The values returned by *myVCCmodel* are pressure, specific enthalpy, temperature, and specific entropy. These values can be used to visualize the thermodynamic cycle in a pressure versus specific enthalpy (hereafter called P-h) plot and a temperature versus specific entropy (hereafter called T-s) plot. The code developed in this work to create these plots make use of CoolProp's *Plots* module and its *PropertyPlot* definition in order to add the thermophysical properties of the refrigerant to the plot. The definitions *myPhPlot* and *myTsPlot* developed in this current work are not shown here for brevity, however they can be found in the online repository in the utilities folder.

It is interesting to visualize the cycle in the P-h and T-s diagrams, particularly in terms of the cycle's position (or proximity to the critical point at the top of the saturation envelope), the degree of temperature lift between the process lines through the evaporator and condenser, and the degree of entropy change in the process line through the compressor (i.e., the slope). Besides these notable cycle characteristics, it is important to quantify the performance of the cycle according to the COP (coefficient of performance). The COP is defined as the ratio between useful and spent energy, or $\dot{Q}_h/\dot{W}_{comp}$ in the case of heating mode, which is reduced to a formula based on only specific enthalpy differences:

```
def myVCCmodel(Tevap, Tcond, SH, SC, n, fluid):

    # input checking
    if  n <= 0 or n > 1:
        print('Problem:')
        print('n should be specified 0 < n =< 1.')
        print('Check calculations.')
        return

    if SC < 0:
        print('Problem: can not specify negative value.')
        print('Check calculations.')
        return

    # Point *1*: inlet to compressor
    T1 = Tevap + SH
    P1 = PropsSI("P", "T", T1, "Q", 1, fluid)
    H1 = PropsSI("H", "P|gas", P1, "T", T1, fluid)
    S1 = PropsSI("S", "H", H1, "P", P1, fluid)

    # Point *3*: inlet to strupeventil
    T3 = Tcond - SC
    P3 = PropsSI("P", "T", Tcond, "Q", 0, fluid)

    # Point *2*: inlet to cond
    P2 = P3
    H2_is = PropsSI("H", "P", P2, "S", S1, fluid)
    H2 = (H2_is - H1) / n + H1
    T2 = PropsSI("T", "H", H2, "P", P2, fluid)
    S2 = PropsSI("S", "P", P2, "H", H2, fluid)

    # Point *3* again
    H3 = PropsSI("H", "T|liquid", T3, "P", P3, fluid)
    S3 = PropsSI("S", "H", H3, "P", P3, fluid)

    # Point *4*: inlet to evap
    H4 = H3
    T4 = Tevap
    P4 = P1
    S4 = PropsSI("S", "H", H4, "P", P4, fluid)

    # Fill out some other points:
    # sat. vap point between points *2* and *3*
    T23_v = Tcond
    H23_v = PropsSI("H", "P", P2, "Q", 1, fluid)
    S23_v = PropsSI("S", "P", H2, "Q", 1, fluid)
    P23_v = P2
    # sat. liquid point between points *2* and *3*
    T23_l = Tcond
    H23_l = PropsSI("H", "P", P2, "Q", 0, fluid)
    S23_l = PropsSI("S", "P", P2, "Q", 0, fluid)
    P23_l = P2
    # sat. vap point between points *4* and *1*
    T41_v = Tevap
    H41_v = PropsSI("H", "P", P1, "Q", 1, fluid)
    S41_v = PropsSI("S", "P", P1, "Q", 1, fluid)
    P41_v = P1

    # Samler variablene og returnerer:
    T = [T1, T2, T23_v, T23_l, T3, T4, T41_v, T1]
    S = [S1, S2, S23_v, S23_l, S3, S4, S41_v, S1]
    P = [P1, P2, P23_v, P23_l, P3, P4, P41_v, P1]
    H = [H1, H2, H23_v, H23_l, H3, H4, H41_v, H1]

    return P, H, T, S
```

Figure 1: Entire code used in the *myVCCmodel* definition.

$$COP = \frac{\Delta h_{cond}}{\Delta h_{comp}} = \frac{h_2 - h_3}{h_2 - h_1} \quad (7)$$

This is because $\dot{Q}_{cond} = \dot{m}\Delta h_{cond}$ and $\dot{W}_{comp} = \dot{m}\Delta h_{comp}$ as previously presented.

The amount of heat transfer from the hot-fluid stream (i.e., the fluid releasing heat energy) to the cold-fluid stream (i.e., the fluid receiving heat energy) can be expressed according to the log-mean temperature difference (LMTD) across the inlets and outlets of the heat exchanger:

$$\dot{Q}_{cond} = U_{cond} A_{cond} LMTD_{cond} \quad (8)$$

and

$$\dot{Q}_{evap} = U_{evap} A_{evap} LMTD_{evap} \quad (9)$$

where U and A are the overall convective heat transfer coefficient and the surface area between fluids, respectively. When calculating the LMTD, a double-pipe heat exchanger is assumed. As such, the equation implemented in the *getMyLMTD* definition is:

$$LMTD = \frac{\Delta T_1 - \Delta T_2}{ln\,(\Delta T_1 / \Delta T_2)} \quad (10)$$

where ΔT_1 and ΔT_2 are the differences between the hot and cold fluid streams between the inlet and outlet of the heat exchanger, depending on whether the fluid streams are parallel or counter-flow. The results in Example 3 are based on a counter-flow configuration, but the code is set up to handle parallel-flow as well. While a double-pipe heat exchanger is assumed in this work, there is of course a possibility to extend the code to include other heat exchanger configurations. For example, a shell-and-tube or a plate heat exchanger can be implemented, as was done and presented in Svortevik (2023).

The *myVCCmodel* requires a value for compressor efficiency. In Example 1, the value for n_{comp} is assumed. However, in Example 2, the value for n_{comp} is calculated in the compressor model based on the pressure ratio between the condenser and evaporator. The compressor model used comes from Corberan et al., (2000) and was also used in Ouadha et al., (2008). It is based on empirical data specific to a compressor type, and takes the polynomial form of:

$$n_{comp}(PR) = A + B * PR + C * PR^2 \quad (11)$$

where the coefficients A, B, and C are 0.66768, 0.0025, and -0.00303, respectively. While this particular compressor model is used here, other compressor models could be easily implemented in this work's online repository as explained in Example 2 in the Results section.

Example 3 was inspired by Camdali et al., (2015) where the cycle is found iteratively such that it yields a predefined heating capacity, $\dot{Q}_{cond}$, while fitting into predefined operating conditions given by Tcond,in, Tcond,out, Tevap,in, and Tevap,out. Similar work was presented in Svortevik (2023), and results presented here in this work is based on the online version of the code. The iterative approach taken in Camdali et al., (2015) and implemented in here in this work could be viewed as a brute-force optimization approach, and that other optimization methods could be applied to solve for the cycle that meets the required heating capacity while minimizing discrepancies between the modelling equations.

3. Results

3.1. Example 1

In the first example, values for Tcond, Tevap, and compressor efficiency are considered to be known input parameters, as well as refrigerant name, amount of super-heating, and amount of sub-cooling; see Table 1. The VCC-calculator (named *myVCCmodel* in Figure 2, code line 17) uses these input values to get P, h, T, and s at the 7 main state points. The cycle is then drawn in a P-h and T-s diagram (see code lines 18 and 19 which call *myPhPlot* and *myTsPlot* respectively). Figure 4 shows the P-h and T-s diagrams for the cases considered in this example. Cycle performance *can* be quantified by the COP, calculated via Eqn. 7 which is based solely on the specific enthalpies across the condenser and compressor, however is not presented here since this example is meant to be illustratively rather than for design or operation purposes. Mass flow rate *cannot* be calculated until either $\dot{Q}_{cond}$, $\dot{Q}_{evap}$, or $\dot{W}_{comp}$ are specified, and is thus not presented here.

Table 1: Case numbers and their input values considered in Example 1.

Case	Tevap (K)	Tcond (K)	n_{comp}	SH (K)	SC (K)	fluid
1-1	270	300	1	0	0	R134a
1-2	270	300	0.6	5	5	R134a
1-3	265	325	0.6	5	5	R134a

```
8  from utils.myVCCmodels import myVCCmodel
9  from utils.myPlots import myPhPlot, myTsPlot
10
11
12 if __name__ == "__main__":
13
14
15     # example 1: show basic functionality
16     fluid = 'R134a'
17     P, H, T, S = myVCCmodel(280, 300, 5, 5, 1, fluid)
18     myPhPlot(P, H, fluid)
19     myTsPlot(T, S, fluid)
```

Figure 2: Snippet of Python code used to run Example 1.

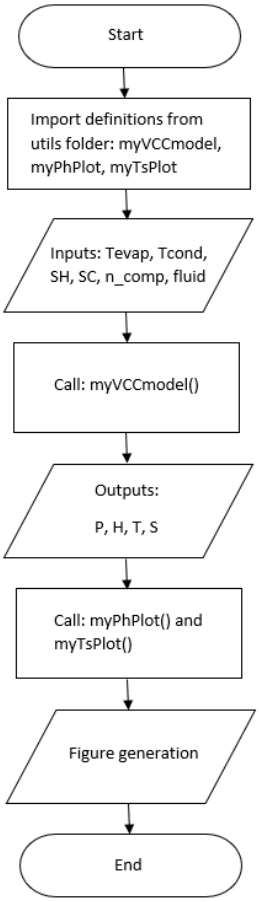


Figure 3: Flowchart of algorithm used in Example 1.

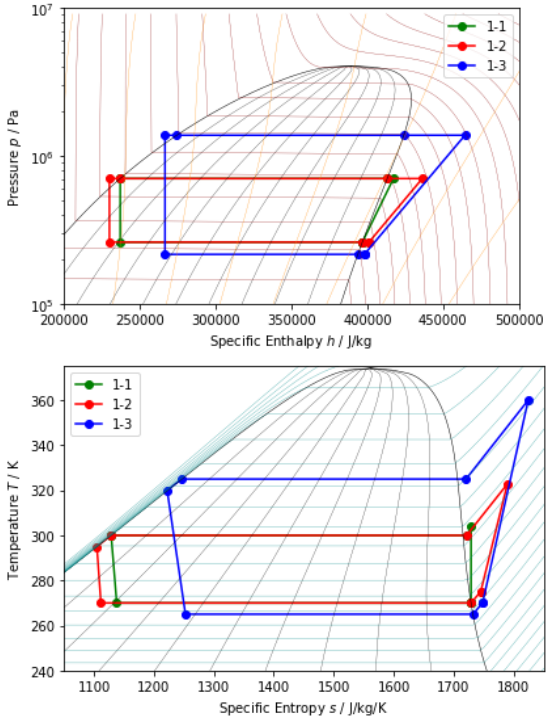


Figure 4: Cycle results for Example 1 obtained using this work's VCC-calculator, illustrated in P-h (top) and T-s (bottom) diagrams.

3.2. Example 2

The second example builds off the first in that Tcond and Tevap are considered to be known (as well as SH, SC, and fluid), however compressor efficiency is given as a function of the pressure ratio. This is more realistic than treating compressor efficiency as a fixed value. Figure 5 shows the code for this example, where code line 22 calls *myCompressor1* which is a definition contained in the *myCompressorModels* module. The compressor efficiency model used here comes from Ouadha et al., (2008), however other compressor models could be easily implemented in the online repository by first creating a copy of *myCompressor1* definition, renaming it *myCompressor2* or something else suitable, and updating the empirically-based expressions for n_vol and n_comp. Table 2 summarizes the cases studied for Example 2. Table 3 presents the pressure ratios that correspond to the Tcond and Tevap input values, and then the compressor efficiency values. Once the compressor efficiency is determined, all of the 7 state points are determined by the VCC-calculator (code line 23 in Figure 5), and P-h and T-s diagrams of the cycle can be made (code lines 24 and 25, respectively). Again, mass flow rate is undefined until $\dot{Q}_{cond}$, $\dot{Q}_{evap}$, or $\dot{W}_{comp}$ are specified. Cycle results are shown in Figure 8, and once again the focus is illustratively rather than conclusive.

```
9  from utils.myVCCmodels import myVCCmodel, getMyPR
10 from utils.myPlots import myPhPlot, myTsPlot
11 from utils.myCompressorModels import myCompressor1
12
13
14 if __name__ == "__main__":
15
16
17     # example 2: show how to include compressor data
18     fluid = 'R134a'
19     Tevap = 280
20     Tcond = 300
21     myPR = getMyPR(Tevap, Tcond, fluid)
22     n_vol, n_comp = myCompressor1(myPR)
23     P, H, T, S = myVCCmodel(Tevap, Tcond, 5, 5, n_comp, fluid)
24     myPhPlot(P, H, fluid)
25     myTsPlot(T, S, fluid)
```

Figure 5: Snippet of Python code used to run Example 2.

```
9  def myCompressor1(PR):
10
11     # compressor specs given in:
12     # Ouadha et al., 2008
13     # (https://doi.org/10.1504/IJEX.2008.019115)
14
15     # input checking
16     if PR < 1.5:
17         print('Problem: this compressor model is not')
18         print('defined for your specificed PR.')
19         return
20
21     n_vol = 1.95125 - 0.80946*PR + 0.17054*PR**2 - 0.01221*PR**3;
22     n_comp = 0.66768 + 0.0025*PR - 0.00303*PR**2
23
24     return n_vol, n_comp
```

Figure 6: Python-definition of *myCompressor1* used to calculate compressor efficiency in Example 2.

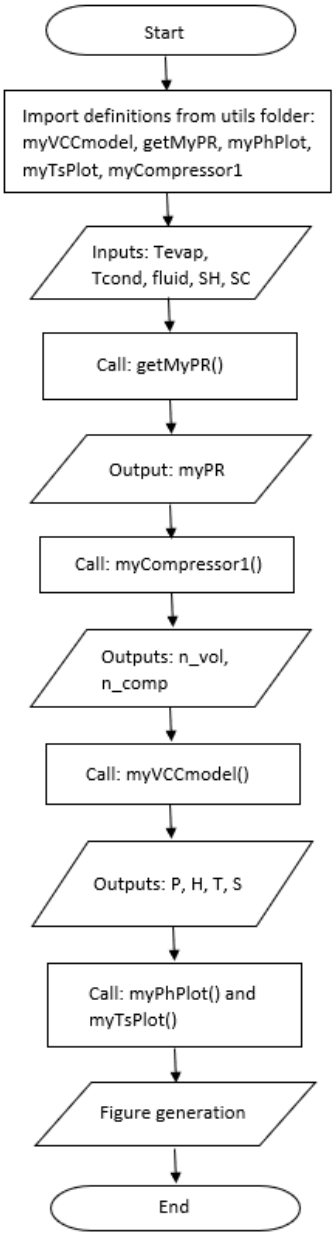


Figure 7: Flowchart of algorithm used in Example 2.

Table 2: Case numbers and their input values considered in Example 2.

Case	Tevap (K)	Tcond (K)	SH (K)	SC (K)	fluid
2-1	270	300	0	0	R134a
2-2	270	300	5	5	R134a
2-3	265	325	5	5	R134a

Table 3: Corresponding pressure ratios and calculated compressor efficiencies.

Case	Pcond (kPa)	Pevap (kPa)	PR	n_{comp}
2-1	702.8	260.8	2.69	0.6524
2-2	702.8	260.8	2.69	0.6524
2-3	1380.3	215.7	6.40	0.5596

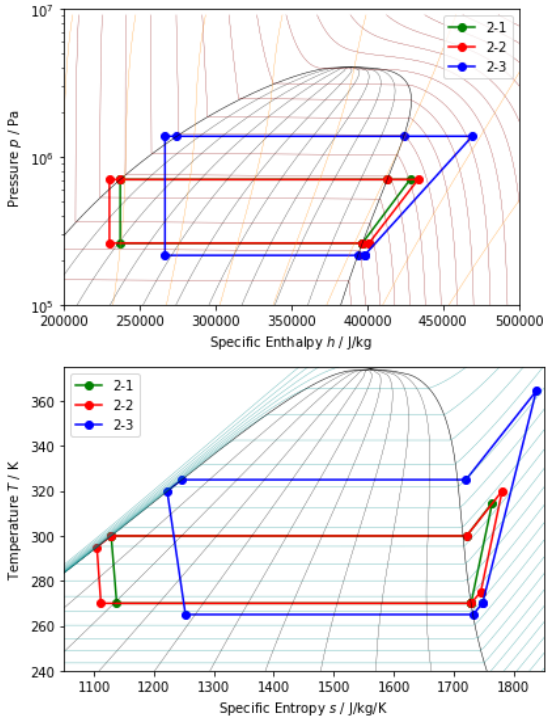


Figure 8: Cycle results for example 2 obtained using this work's VCC-calculator, illustrated in P-h (top) and T-s (bottom) diagrams.

3.3. Example 3

The last example aims to be a bit more practical than Examples 1 and 2. That is, it is unlikely that Tcond and Tevap are known, but rather, an amount of heat output from the heat pump cycle, $\dot{Q}_{cond}$, is more likely to be specified, along with values for Tcond,in, Tcond,out, Tevap,in, and Tevap,out. The required $\dot{Q}_{cond}$ can be used in combination with Tcond,in and Tcond,out values to calculate a Tcond value. But the rest of the cycle must be such that it agrees with this Tcond. In other words, the question is what values do Tcond and Tevap take on such that the cycle yields the required until $\dot{Q}_{cond}$? This question can be answered using an iterative approach that is similar to what was presented in Camdali et al., 2015.

Cases are presented in Table 4, where SH, SC, and fluid are 5, 5, and R134a, respectively, and compressor efficiency is given as a function of pressure ratio by Eqn. 11. Also, U and A values for both the condenser and evaporator are set to 1000 W/m^{2}K and 2 m^2, respectively. Results are shown in Table 5 and Figure 11. This time, mass flow rate of the refrigerant can be determined since $\dot{Q}_{cond}$ is specified. Figure 12 shows the convergence behavior of two main quantities or residuals, namely how close the cycle's Qcond value (black x's) is from the wanted heat capacity (red dashes),

and how close Qevap1 and Qevap2 are from each other (blue dots) since their difference should be as close to zero as possible. The cycle was found after about 2300 iterations. Other methods could be used to obtain this cycle solution faster, for example, using an optimization method where the objective is to minimize the residuals.

Table 4: Case numbers and their input values considered in Example 3.

Case	Capacity $\dot{Q}_{cond}$ (kW)	Tcond,in / Tcond,out (°C)	Tevap,in / Tevap,out (°C)
3-1	30	30/40	8/0

Table 5: Results for cases considered in Example 3.

Case	#its. needed	Tcond	Tevap	n_{comp}
3-1	2349	323.55	255.50	0.4411

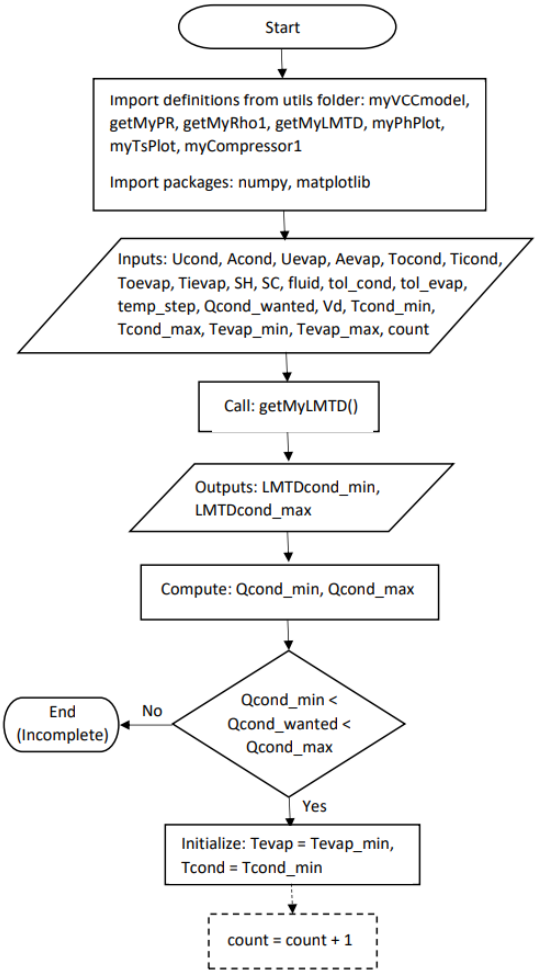


Figure 9: First half of flowchart for Example 3.

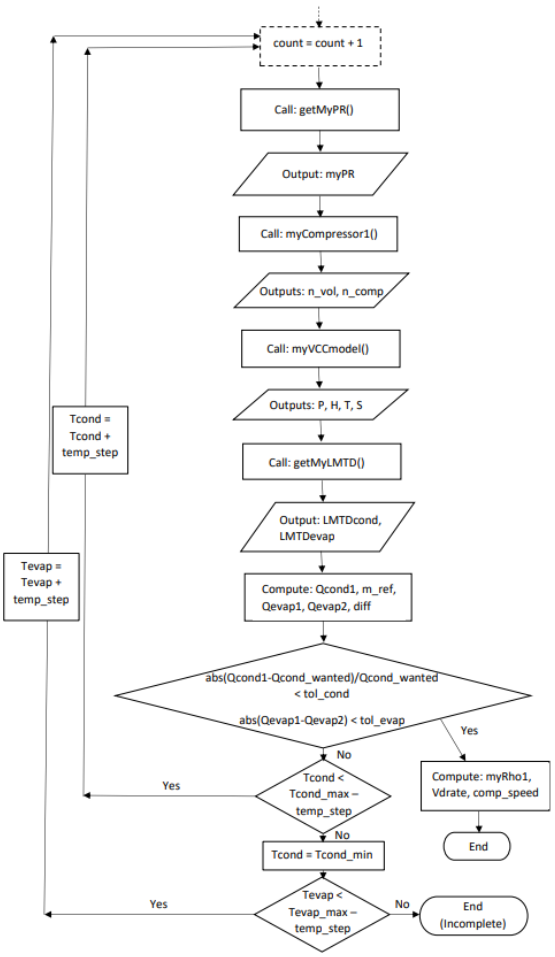


Figure 10: Second half of flowchart for Example 3.

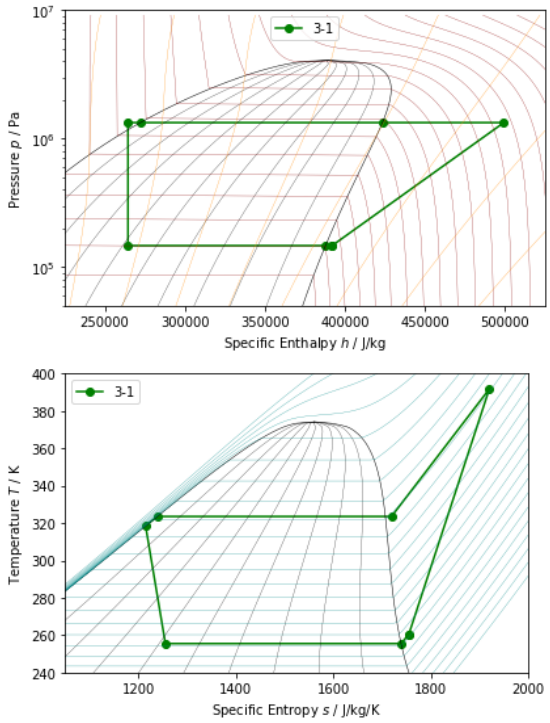


Figure 11: Cycle results for example 3 obtained using this work's VCC-calculator, illustrated in P-h (top) and T-s (bottom) diagrams.

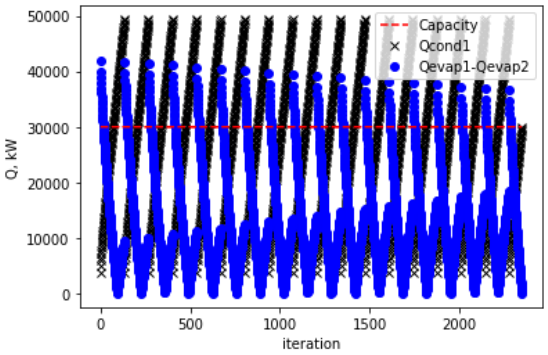


Figure 12: Convergence of solution towards a cycle that yields the wanted heat capacity with a minimal discrepancy between Qevap1 and Qevap2.

4. Summary

This work has presented the typically used modelling equations to simulate the thermodynamics of the vapor compression cycle. Three examples were presented to demonstrate the structure and capability of the code. A main contribution is the ability to implement various compressor models that can come from empirically derived expressions. Another contribution is the implementation of a similar iteration scheme presented by Camdali et al., 2015, however in Python versus MATLAB, which means the code can be assessed by more researchers, engineers, or others in the heat pump community. The method taken in Example 3 is similar to a brute-force optimization approach and other optimization methods could be implemented instead to reduce computational intensity.

The purpose of the three examples was illustrative in nature rather than to produce values that represent an engineering design or operation point of a heat pump. Indeed, this work marks the start of a developing online code that has the potential for modification and extension, and thus the focus here was to show basic structure of the VCC-calculator and how it can be used. As such, this paper did not focus too much on interpreting results, as they are intended to be illustrative only.

Source code used in this work can be found online: https://github.com/AllenGitCode/VCCmodelling.

References

Aulicino, C., and S. Bakrania, S., (2022) 'A Python-based lab module to conduct thermodynamic cycle analysis,' *IEEE Frontiers in Education Conference (FIE), Uppsala, Sweden, 2022*, pp. 1–6, doi: 10.1109/FIE56618.2022.9962388

Bell, I.H., Wronski, J., Quoilin, S., and Lemort, V. (2014) 'Pure and Pseudo-pure Fluid Thermophysical Property Evaluation and the Open-Source Thermophysical Property Library CoolProp,' *Industrial and Engineering Chemistry Research*, 53, pp. 2498–2508. doi: 10.1021/ie4033999

Camdali, U., Bulut, M., and Sozbir, N., (2015) 'Numerical modeling of a ground source heat pump: The Bolu case,' *Renewable Energy*, Elsevier, 83(C), pp. 352–361. doi: 10.1016/j.renene.2015.04.030

Corberan, J.M., Urchueguia, J., Gonzalves, J. and Calas, A. (2000) 'Performance of a reciprocating hermetic refrigerant compressor using propane as working fluid', *Proceeding of the 4th IIR-Gustav Lorentzen Conference on Natural Working Fluids at Purdue*, IIF-IIR Commission B1, B2, E1, E2, pp.225–232.

IEA HPT. 'HTP – Heat Pumping Technologies'. https://heatpumpingtechnologies.org/ Accessed on Aug 17, 2023.

Jensen, J. K., Ommen, T., Reinholdt, L., Markussen, W. B., & Elmegaard, B. (2018) 'Heat Pump COP, part 2: Generalized COP estimation of heat pump processes.' In Proceedings of the 13th IIR-Gustav Larentzen Conference on Natural Refrigerants (Vol 2, pp. 1136-1145). International Institute of Refrigeration. https://doi.org/10.18462/iir.gl.2018.1386

Madessa, H. B., Torger, B., Bye, P. F., Erlend, A., (2017) 'Parametric study of a vertically configured ground source heat pump system.' *Energy Procedia*, 111, pp. 1040-1049. doi: 10.1016/j.egypro.2017.03.267

Ouadha, A., En nacer, M., and Imine, O., (2008) 'Thermodynamic modelling of a water-to-water heat pump using propane as refrigerant,' *International Journal of Exergy*, 5(4), pp. 451–469. doi: 10.1504/IJEX.2008.019115

Svortevik, E., (2023) 'Python-based modeling of the vapor compression cycle focusing on heat exchangers and user-friendliness with web page,' Master's thesis, Oslo Metropolitan University.

Vering, C., Engelpracht, M., Göbel, S., Hoseinpoori, S., Wüllhorst, F., Schwenzer, C., Rademacher, M., Hinrichs, S., Chandra, F., Mehrfeld, P., and Müller, D., (2022) 'Open-Source vapor compression library (VCLib): Heat pump modeling for education and research,' *Comput. Appl. Eng. Educ*. 30, pp. 1498–1509. doi: 10.1002/cae.22540

Wang, J., Qv, D., Ni, L., Fan, J., Kong, W., (2022) 'Matching-design for inverster air-source heat pump system based on heating load characteristics of civil buildings,' *Energy and Buildings*, 260(111952). doi: 10.1016/j.enbuild.2022.111952

Implementation of a bolted joint model in Modelica

Nils Dressler[a,b,*] Lars Eriksson[b]

[a] *Atlas Copco Industrial Technique AB, Sweden* [b] *Vehicular Systems , ISY - Linköping University, Sweden*
[*] nils.dressler@atalscopco.com

Abstract

The basic mechanics of a bolted joint are well-known and have been studied for a long time. The dominating principle is to represent the parts in a joint as a series connection of linear compression and tension springs. However, traditional models often neglect the tightening dynamics and their interrelation with, for instance the friction or embedment. To study these phenomena further and determine their impact on the tightening process and dynamics, and for developing new tightening control strategies, it is necessary to model a threaded fastener and implement it in a suitable simulation environment.
Existing models and experimental data have been studied to find equations that fit the observed behavior. Novel models were combined with standard Modelica components to form a threaded fastener model. The simulation results were compared with tightening data from experiments. This work proposes new models for the first three tightening phases, embedment, and threaded fastener friction. These models are implemented in the modeling language Modelica. The results show that it is possible to resemble a typical threaded fastener tightening with power tools. The friction and tightening phases show the expected behavior, while the embedment model needs further experimental verification. During modeling, the model is susceptible to the chosen parameters. Parameters for the joint stiffness, obtained via the VDI guidelines, needed to be reduced by 30% to resemble the joint in a dynamic simulation.

1 Introduction

Threaded fasteners are often referred to as the most common machine element, and therefore, the importance of threaded fastener joint reliability is apparent. The generated clamp force is difficult to measure but of great importance for the functionality. Therefore, it is of great interest to study the behavior of threaded fasteners under dynamic tightening conditions, and the approach of modeling the fastener is a first step for validating theories, e.g., the impact of friction behavior.

2 System Overview

A threaded fastener joins or holds together two or more components or materials. This is done via the clamp force.
A threaded fastener assembly typically involves an operator, a power tool, and a threaded fastener. This work focuses on threaded fastener behavior.
In its simplest form, a joint is composed of a bolt, a nut, and at least two clamped parts. Additional parts like washers or gaskets are common. In a tightened joint, the bolt is under tension between the bolt head and the engaged thread, while the clamped parts are under compression. Friction occurs during tightening under the fastener head and the adjacent surface and between the bolt threads and the nut. Usually, up to 90% of the applied torque in a tightening goes to overcoming the friction. This highlights the importance of understanding friction during tightening. A difference from many other systems subjected to friction is that the normal load and friction torque are constantly increasing. This leads to very high friction torques and explains the large share of friction losses on the total energy put in the system.
Power tools used for tightening are typically composed of an electrical or pneumatic motor, gears, drive shafts, and housing.

2.1 Tightening Mechanics

The tightening mechanics can be separated into two domains: the rotational domain with driving and load torques and the translational domain with the preload and clamp force.

2.1.1 Rotational Domain

The Kellerman-Klein equation adapted by (VDI - Verein Deutscher Ingenieure, 2015) and originating from (Kellermann & Klein, 1956) describes the rotational domain (see Equation 1). The formula is derived from a special case of Newton's second law where the angular acceleration is zero. M_t is the driving torque, and F_c is the clamp force. The origin of that equation can better be understood by studying the free

body diagram of the bolt thread; this can be found in the chapter Torque and Tension in Fasteners in (Oberg et al., 2004).

$$M_t = F_c \left(\frac{P}{2\pi} + 0.58 d_2 \mu_{th} + \frac{D_b}{2} \mu_h \right) \quad (1)$$

The under-head friction torque is the product of the under-head friction radius $\frac{D_b}{2}$, the clamp force, and the underhead friction coefficient μ_h. The pitch torque is the product of the clamp force and the thread pitch P. The thread friction torque is the product of the thread mean radius $\frac{d_2}{2}$, the thread friction coefficient μ_{th}, and lumped geometric parameters.
A model for that domain is found in (Japing et al., 2015).
The transformation from pitch torque into a linear force is done via the thread.

2.1.2 Translational Domain

In the translational domain, two forces act: the preload and the clamp force. The force balance is often visualized with joint diagrams. Such a joint diagram can be seen in more detail in (Shoberg, 2000). The bolt elongation is usually larger than the joint compression, but due to different resilience, the forces are equal. A typically used analogy for the interaction between the bolt and the clamped parts is a two-spring model where a tension and a compression force are coupled in series. During tightening, the fastener constantly rotates relative to the nut. Every angular change $\Delta\varphi$ causes the nut to move upwards on the bolt thread, called Δs. Equation 2 describes the upward movement.

$$\Delta s = \Delta\varphi \cdot P \quad (2)$$

The resulting force increment is described via Equation 3, where s is the length and c is the stiffness. As seen, the total force equals the force in the bolt and joint.

$$\Delta F_c = \Delta s \cdot c_{\text{tot}} = \Delta s_{\text{bolt}} \cdot c_{\text{bolt}} = \Delta s_{\text{joint}} \cdot c_{\text{joint}} \quad (3)$$

The mechanics in more detail can be found in literature (Toth, 2006).

2.2 Tightening Phases

A threaded fastener tightening is usually divided into four different tightening phases. Run Down, Alignment, Linear Elastic Clamping, and Yield. The classification is done via torque traces in the angle or time domain, as in Figure 1
The phases are characterized by the following.
Run Down: The nut is not yet in touch with the clamped parts during run down. The resistance to overcome is friction in the thread due to interference and the acceleration of the rotating masses. The torque during run-down is assumed to be constant. The run-down ends when the torque increases from that constant level. The clamp force is zero during run down.

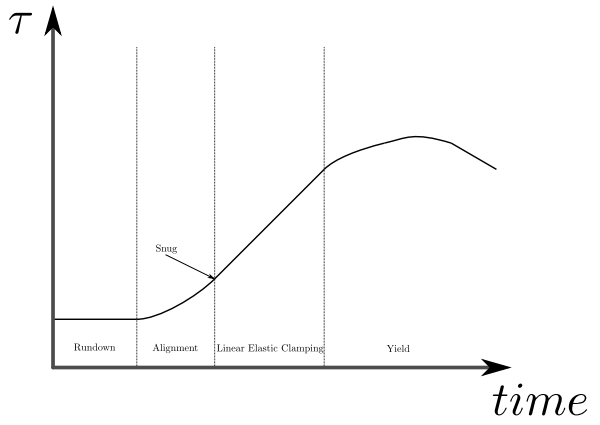


Figure 1. Torque evolution during the tightening phases of a constant speed tightening

Alignment: The alignment phase begins when the nut and clamped parts come into contact. The tightening torque increases at a larger non-constant rate. Reasons for the non-linearity, material imperfections, and the initial alignment of the joint components. Alignment ends at snug when the torque increment transitions to a constant rate. The clamp force build-up rate is not constant in this phase.
Linear elastic clamping: In the linear elastic clamping phase, the torque and preload increase happen at a constant rate. Many standard tightening methods, based on torque or angle measurements, end the tightening in that phase.
Yield: Yield is the last phase of the tightening. It starts when the material behavior of the fastener shaft changes from linear elastic to plastic deformation. In the yield phase of a tightening, a linear relationship between the tightening torque and the tightening angle ends.

2.3 Embedment

Embedment during and after tightening leads to a clamp force loss over time. Most prominent is the clamp force loss due to embedment that can be observed after a finished tightening or at short resting times or pauses during tightening. Embedment occurs due to high local stresses on rough surfaces. These high stresses lead to local plastic deformation in the contact regions. As a result, the total length of a component is shortened, which leads to a clamp force loss. Experimental data shows that the clamp force loss rate decays over time. To account for the clamp force loss, fixed losses are assumed based on the surface roughness of the joint components. The exact dynamics of embedment are not analytically described. It can be concluded that a longer tightening duration results in

less post-target embedment. Another insight is that more embedment happens at higher loads.

3 Modelica Implementation

The previously described system is the basic structure combining a rotational oscillator and a translational oscillator. The connection between these two oscillators is via the kinematics in the thread, which converts between the rotational and translational domains. This system is driven by an input torque and is overdamped due to friction. Implementing such a complex system with a component-based modeling language like Modelica allows rapid model development and offers a convenient way of evaluating different submodels.
The entire model is built mainly with components from the Modelica standard library. Components from the mechanics rotational and translational libraries are used. The components developed explicitly for the threaded fastener model are: *HeadFriction*, *ThreadFriction*, *IdealThread*, *ThreePhaseBolt*, and *Embedment*.
The composed model can be seen in Figure 2.
The order of the models in the rotational domain is the following from left to right: *Control block*, *Torque*, *TorqueSensor*, *BearingFriction*, *Inertia*, *Spring* and *Damper*, *HeadFriction*, *Inertia*, *Spring* and *Damper*, *ThreadFriction*, *BearingFriction*, *Inertia*.
Connected via the *IdealThread* follows the translational domain, the models are from right to left: *ForceSensor*, *ThreePhaseBolt*, *Mass*, *Embedment*, *SpringDamper*, *Fixed*.
The translational domain has one difference between the theoretical model and the implementation. The joint and the bolt are modeled as compression springs, contrasting the tension and compression spring serial connection found in the literature.

3.1 Friction Models

The implementations of the friction models can be seen in Appendix A.1 for the head friction model. Apart from the two parameters, the thread friction model is identical and therefore not shown. Both are based on the rotational Brake model from the Modelica standard library.
The friction implementation is a state machine with the states: *backward*, *forward*, *free*, or *stuck*. The transition to the stuck state is made when the velocity is zero. The acceleration of the component is set to zero in the stuck state. The condition for a forward and backward movement is that the sum of the external torques is larger than the defined friction torque at zero velocity, multiplied by the given peak factor. This behavior is essential for the friction component to behave like friction in threaded fasteners. Otherwise, the fastener would unwind after the input torque is removed.
The implementation is coulomb friction combined with speed-dependent dynamic friction. For the fastener model, a friction coefficient that increases with speed is essential to avoid oscillations of the inertia and masses.
The remaining adaptations are to align the parameters with the friction radii as in Equation 1. For the thread friction model, d_2 is directly given as a parameter. For the head friction model, the friction radius $\frac{D_b}{2}$ is calculated from the plane head bearing diameter of the bolt d_W and the plane head bearing areas inside diameter D_{ki}.

3.2 Embedment Model

The embedment model is based on the rod model from the Modelica standard library. The difference is that the component has a variable length that becomes negative under pressure. The model can be found in Appendix A.2.
The length change is described by Equation 4. The authors invented this equation to model the clamp force loss due to embedment.

$$\dot{L} = -\frac{1}{\tau}\left(L - L_{\max}\left(\frac{f}{f_{\max}}\right)\right) \tag{4}$$

Here, L is the length of the embedment surface, $L_{\max}$ the maximum possible embedment, f the force applied, $f_{\max}$ the maximum possible force, and τ the time constant defining how fast the process happens.

3.3 Thread Model

The thread model is based on the *IdealGearR2T* model from the Modelica standard library. The functionality is identical, and only the input parameters have been changed to align with the terminology from threaded fasteners. The model can be seen in Appendix A.3.
The model parameter is the pitch per revolution P instead. The torque (τ) and force (f) relationship is given by Equation 5. The rotation (φ) displacement (s) relationship is given by Equation 6.

$$\varphi\frac{P}{2\pi} = s \tag{5}$$

$$\tau = f\frac{P}{2\pi} \tag{6}$$

3.4 Tightening Phases

The first three tightening phases are distinguished by what happens with the joint and bolt.
During run down, the nut and the joint are not yet in contact, so neither the bolt is stretched nor the joint compressed. Since the rotation is applied to the bolt,

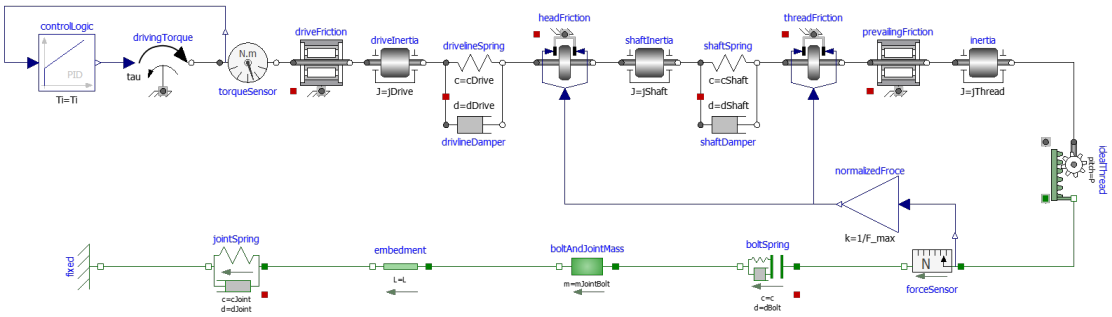


Figure 2. Threaded fastener model, implemented with Open Modelica

the bolt can be seen as a rotating inertia at a free end. During the alignment phase, when the nut and the joint come in contact, the bolt is stretched while the joint is compressed. The displacement force relationship, and thereby even the displacement-torque relationship, is not linear. The rotation of the bolt is opposed by a resistance that is ultimately caused by the bolt stretch and joint compression. Therefore, the rotating bolt can no longer be considered inertia on a free end.
In the linear-elastic clamping phase, the displacement-force relationship changes from non-linear to linear. These effects are lumped into one component. The choice was made to lump the free-end and non-linear behavior into the bolt model, creating the three-phase bolt. Further experiments would be needed to identify what component contributes how much to the nonlinearities, and due to the lack of insight, these effects are combined into the bolt with the reasoning that the bolt is earlier in the drive line and that this arrangement will minimize the risk for unintended oscillations.

The model for the three-phase bolt is based on the *ElastoGap* component from the Modelica standard library. The spring force for the three-phase bolt is modeled according to Equation 7.

$$f_c = \begin{cases} 0 & \text{if} \quad s_{\text{rel}} < s_{\text{rel0}} \\ c_{\text{quad}} |s_{\text{rel}} - s_{\text{rel0}}|^2 & \text{if} \quad s_{\text{rel0}} \geq s_{\text{rel}} < s_{\text{rel2}} \\ c_{\text{lin}} |s_{\text{rel}} - s_{\text{rel1}}| & \text{else} \end{cases} \tag{7}$$

For such a model, c_{lin}, s_{rel0} $s_{\text{rel}p2}$ are the model parameters for the linear spring constant, the rundown displacement, and the relative displacement for the second phase, namely the alignment. The remaining parameters c_{quad}, s_{rel1} are the quadratic spring constant and the hypothetical crossing of the linear spring phase with the zero force line.

The parameters c_{quad}, s_{rel1} can then be derived through the fact that there is a smooth transition between the phases, which means that the derivative is equal in the transition points. That yields Equations 8, 9, and 10 for the parameters.

$$s_{\text{rel}_1} = s_{\text{rel2}} - \frac{s_{\text{rel2}} - s_{\text{rel0}}}{2} \tag{8}$$

$$s_{\text{rel}_2} = s_{\text{rel0}} + s_{\text{rel}p2} \tag{9}$$

$$c_{\text{quad}} = \frac{c_{\text{lin}}}{2(s_{\text{rel2}} - s_{\text{rel0}})} \tag{10}$$

The model for the implementation is given in Appendix A.4

4 Verification

The objective of the fastener model is to approximate a real tightening trace from tightening experiments. For that, the two-step tightening from (Persson et al., 2021) is taken as a reference. The speed profile, the clamp force trace, and the torque trace can be seen in Figures 3, 4, and 5, respectively. Here, the data from the experiments is marked as recorded data in blue. The joint is an M10×70 hexagonal flange head fastener type of strength class 8.8 with Zn-Fe coating + wax. The clamp length of the joint is 56 mm, and the coefficient of friction has been determined to be 0.147 ± 0.016 (±3σ) according to ISO16047 at a tightening speed of 20 rpm. The set target torque was 43 Nm.
An initial guess for the simulation parameters is made with the given data. They are based on the VDI guidelines ((VDI - Verein Deutscher Ingenieure, 2015)), taken from data sheets of the used equipment, read from the given plots, or are empirical values based on modeling experience.
From the initial guess to the tuned parameters, the Bolt and Joint Stiffness were reduced by 30%. The initial guess has a rundown and alignment angle of 70° and 52°. The remaining parameters remained unchanged.
The tightening parameters are set to 100 rpm until 21 N m are reached. The pause step is 50 ms, and the final step is 20 rpm until 43 N m are reached.
The tuned parameters can be seen in Table 1. The change in % refers to how much the parameter was altered from the initial calculation to the final parameter. Further geometric parameters are d_2 7.19 mm ,

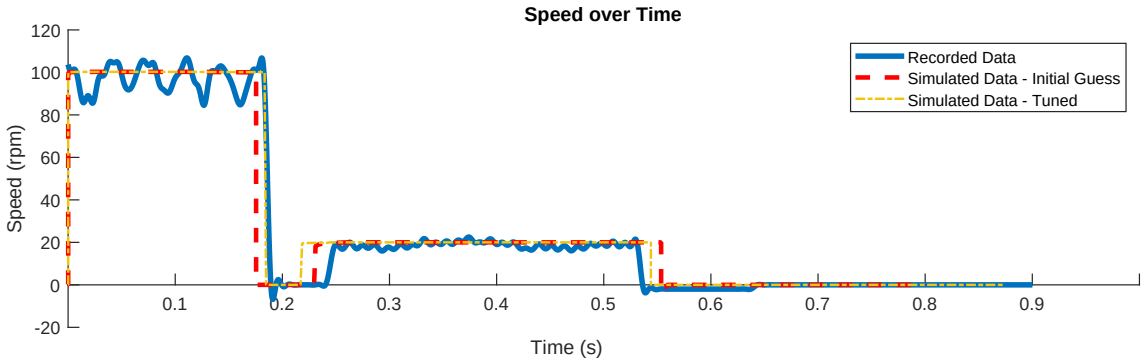


Figure 3. The speed over time trace of the recorded tightening, simulated tightening with initial guess parameters, and the simulated tightening with tuned parameters

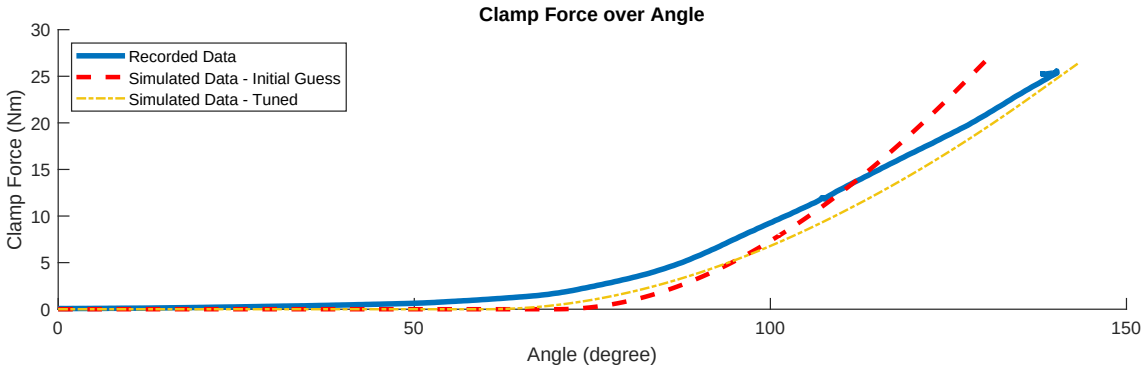


Figure 4. The clamp force over angle trace of the recorded tightening, simulated tightening with initial guess parameters, and the simulated tightening with tuned parameters

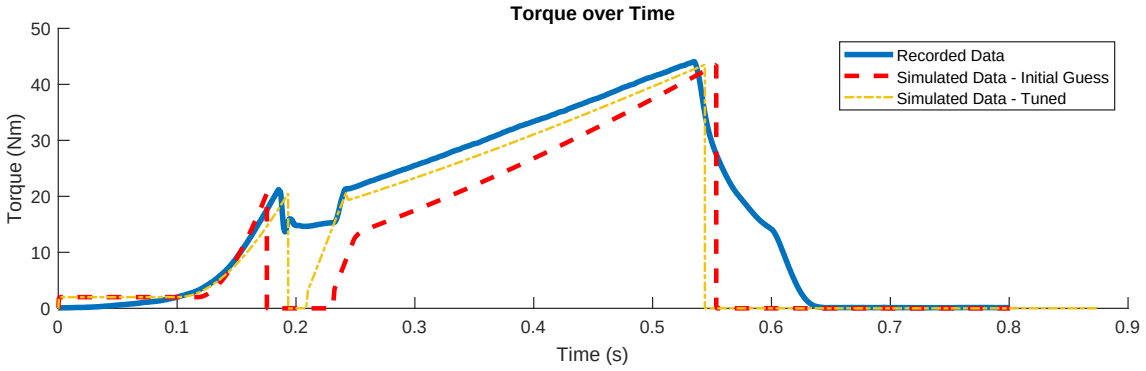


Figure 5. The torque over time trace of the recorded tightening, simulated tightening with initial guess parameters, and the simulated tightening with tuned parameters

Table 1. Tuned parameters used for the simulation

Parameter	Tuned Value	Change in %
Bolt and Joint Mass (kg)	0.3356	0
Bolt Stiffness (N/m)	$1.5758 \cdot 10^8$	30
Bolt Damping (Ns/m)	$2.4106 \cdot 10^2$	0
Shaft Stiffness (Nm/rad)	$1.3685 \cdot 10^3$	0
Shaft Inertia (kgm^2)	$2.8846 \cdot 10^{-7}$	0
Shaft Damping (Ns/m)	$2.4106 \cdot 10^2$	0
Thread Inertia (kgm^2)	$1.7572 \cdot 10^{-7}$	0
Joint Stiffness ((N/m)	$8.6478 \cdot 10^8$	30
Joint Damping (Ns/m)	$1.5279 \cdot 10^3$	0
Driveline Inertia (kgm^2)	$3.45 \cdot 10^{-6}$	0
Driveline Stiffness (Nm/rad)	265.74	0
Driveline Damping (Nms/rad)	$1.63 \cdot 10^{-6}$	0
CoF Thread 0/20 rpm (-)	0.145 / 0.147	-20.8 / 0
CoF Bolt Head 0/20 rpm (-)	0.145 / 0.147	-20.8 / 0
Prevailing Torque(Nm) 0/100 rpm	0.7 / 1.3	0 / 0
Driveline Torque(Nm) 0/100 rpm	0.2 / 0.7	0 / 0
Rundown / AlignmentAngle (°)	60 / 70	-20 / -66
Embedment total (m)	$8 \cdot 10^{-8}$	0
Embedment time constant (s)	0.5	0
Preload at yield (N)	32000	0

D_{Ki} 9 mm , P 1.5 mm, and d_W 11.63 mm. The ratio between static and dynamic friction is 1.5 for all friction components. The control parameters are kp = 100, Ti = 0.005, maximum Torque yMax = 50 Nm, rundown speed = 100 rpm, final speed = 20 rpm, rundown torque = 20.5 Nm, final torque = 43.5 Nm, break = 0.05 s.
As a result of the tuning, a reduction of the bolt and joint stiffness by 30% was made. The length of the alignment angle was increased by 8 degrees, while the rundown angle was shortened by 10 degrees.
The speed profile of the recorded data is only followed to a certain degree, as seen in Figure 3. In the model case, an optimal torque source with no delays. Moreover, a relatively fast controller is used. Due to that follows the modeled result the reference value better. It can nevertheless be seen that the duration of different speeds deviate from the reference.
Overall, the model resembles the tightening data well, even if there is no exact match between the recorded data and the modeled tightening. A better fit can be obtained with further tuning or optimization of the parameters. Regardless of that, it can be seen that the characteristic elements of the tightening are accurately represented with the simulation model.

5 Results and Discussion

5.1 Modelling Process

The modeling process was iterative. The involved components were tested in isolation, in combination with other components, and how different parameter ranges affect the behavior of the components in composed systems. When working with OpenModelica, comparing the simulated results with variations of models and parameter combinations could be more convenient. Tracking and relating the tested parameters to simulation results gets more complicated with a growing system complexity.
A suitable method for that modeling work has been to do the modeling work and system composition within OMEdit and experiment with initial parameters in fast iterations. Then, the model can be loaded into an OMNotebook. There, it is of greater convenience to study how different parameters impact the modeling result while at the same time keeping track of the parameter changes. That is of increased importance for threaded faster modeling due to the interrelation of the parameters, which often prohibits the change of just a single parameter. An example of coupled parameters is the dependency of the mass, stiffness, and damping on the geometry of the component, so a change in length would impact all of them, while a change of one of the mentioned parameters would require the others to be changed to be consistent.

5.2 Model Alignment

The initial parameters obtained via analytical calculations following the VDI2230 guidelines showed discrepancies between the recorded data and the model. This was expected since the calculations are simplifications based on the static case. It can be seen in Figure 4 that the clamp force increases too much with advancement in angle. Similarly, the initial torque trace is rising too fast, as seen in Figure 5. A more similar torque rate was achieved by reducing the bolt and joint stiffness by 30%. The exact reasons for the discrepancy are not further studied but are left for future work. Still, it can be taken as a result that the analytical stiffness for threaded fastener joints overestimates the clamp force rate if applied to the given model.
As seen in Figures 3, 4, and 5, the initial agreement could be more optimal. The model fit can be improved with parameter tuning. Beyond the first shown attempt, this is left to future work.
Overall, the system has a good agreement with the recorded data. Several experiments have been done to verify the model, but only the data from one experiment has been used for tuning. Therefore, the data from further experiments is not included in the results. Non-optimized model parameters can explain the discrepancies and can be minimized further. Hence, the model agrees well with a real threaded fastener joint. One important conclusion is that the data obtained in alignment with the VDI guidelines is either unsuitable for dynamic modeling in general or for the specific implementation of the model.

5.3 Model Components

The modeling work resulted in three new component models and is otherwise composed of standard components from the Modelica library. The models fulfilled

the purpose in the system model and contributed to the overall model agreement with the recorded data.

5.3.1 Friction Components

The purpose of the modified friction components is to resemble the friction under the fastener head and in the thread, respectively. This behavior was already implemented via the brake component from the Modelica standard library. These models are simplified so that most of the effects, such as lubrication and surface profiles, are described through the coefficient of friction. All that was necessary was a redefinition of the model parameters to match bolt nomenclature. Both components work well in the complete system, which can be observed while studying the brake modes. The brake modes indicate in which state the friction component is, which could be either moving backward or forward, stuck or free. The behavior for transitioning from stuck to any other state is essential for the model to resemble the behavior of an actual fastener. This is implemented via the peak ratio. Without that behavior, a threaded fastener would unwind again once the applied torque is taken away or set to zero.
During the development of the model, the friction behavior caused the biggest challenges. It could be observed that the friction components did not transition to a locked state, even though, based on empirical experience from threaded fasteners, they should have. From a simulation perspective, that occurred since the condition to transition to a locked state - a relative velocity of 0 - was never met. That happened especially when the adjacent components had relatively small masses or inertias. One observation is that this behavior does not occur in the overall system with the given parameters. Further studies must be done to determine if a particular parameter combination is causing that behavior. For now, it can only be concluded that in terms of stability, it is more beneficial to handle the entire system than extra single components and test them in separate test scenarios.

5.3.2 Three-phase bolt

The three-phase bolt is a key component in resembling the different tightening phases. As seen in the shown results, the chosen approach works, and the threaded fastener behaves accordingly in the different stages, which are determined via the tightening angle. The difference, especially during the early phases of the tightening, is instead a debate about the presence of a run-down in the recorded data. It can be addressed by shortening the rundown phase and extending the alignment phase in exchange. That could come at the cost of a worse fit and lead to the conclusion that a second-degree polynomial function for the spring force is unsuitable. Further studies have to be done to determine that.
The transition between alignment and the linear elastic phase is smooth and working as in the recorded data; therefore, is the chosen approach sufficient for the modeling purpose.
The recorded data has a relatively short run-down but a much longer alignment phase, as seen in Figure 4. Further studies are needed to confirm that a second-degree polynomial function for the spring force during alignment is a good fit for the case of such a long alignment phase.

5.3.3 Embeddment

The effects of the embedded can be read from the model where a clamp force loss due to a shortened spring length can be detected. In the tightening trace, that can not be detected or distinguished from a tightening with a different spring stiffness. The embedment effect is only vaguely present in the recorded data. A more suitable tightening scenario must be chosen to further investigate the embedment model's alignment. This could be a joint with rougher surfaces to increase the embedment effects and a shorter tightening duration by higher tightening speeds to relocate more of the embedment effect after the final shutdown. For the current model, it can hence only be concluded that it is possible to model embedment in the proposed way, that the effects are present, and that an embedment component in the spring chain does not negatively impact the other components, such as the friction components, as long as the entire threaded fastener model is kept as whole. In test scenarios where the embedment component was tested with only one friction component, one spring, and one inertia, it kept the friction component from locking until the maximum embedment was reached.
Embedment is not very distinct in the recorded data. Therefore, no quantitative comparison is possible. It can be concluded that the chosen embedment model works in principle but that the effects must be studied further.

6 Conclusions

It can be concluded that the proposed model does resemble a threaded fastener during tightening and a simplified driveline. Some differences could be observed. Most of the differences were due to differences in the control input for the system. The remaining deviations can be minimized or reduced by tuning model parameters. Hence, the overall model is a valid representation of a threaded fastener. The main insights generated by the modeling work are:

- The model is very sensitive in terms of parameter combinations

- The parameters obtained by the VDI guidelines overestimate the spring systems stiffness

The combination of these factors impedes the development of fastener models, while the actual parameters are difficult to obtain. At the same time, the model does not work accurately if the wrong model parameters are chosen. Therefore, a good test case with well-known parameters is beneficial when developing models of that type. A comparison with parameters obtained by commercial FEM software and to what model fit they lead could indicate if more suitable stiffness parameters could be obtained for future model development.

Acknowledgment

We thank Erik Persson and his Tightening Technique team from Atlas Copco Industrial Technique for the support and discussions leading to the model's development. Atlas Copco fully finances the work, and therefore, we would like to thank Daniel Brandt for sponsoring the research work. The work was conducted as part of the Linköping Center for Sensor Informatics and Control. Therefore, we would like to thank Svante Gunnarsson and the Ph.D. students within the center for their support.

References

Japing, A., Seibel, A., & Schlattmann, J. (2015). Modellentwicklung zur beschreibungvon reibschwingungen bei der schraubenmontage. *Reibung, Schmierungund Verschleiß: Forschung und praktische Anwendungen. 52. Tribologie-Fachtagung*.

Kellermann, R., & Klein, H.-C. (1956). Berücksichtigung des reibungszustandes bei der bemessung hochwertiger schraubenverbindungen. *Konstruktion*, *8*, 236.

Oberg, Jones, Horton, & Ryffel. (2004). *Machinerys handbook* (27th ed.). New York, NY: Industrial Press.

Persson, E. V., Kumar, M., Friberg, C., & Dressler, N. (2021). *Clamp Force Accuracy in Threaded Fastener Joints Using Different Torque Control Tightening Strategies* (Tech. Rep.). Warrendale, PA. doi: 10.4271/2021-01-5073

Shoberg, R. S. (2000). Engineering fundamentals of threaded fastener design and analysis. i. *Fastening*, *6*(2), 26–29.

Toth, G. (2006). *Torque and angle controlled tightening of bolted joints* (Unpublished doctoral dissertation). Göteborg.

VDI - Verein Deutscher Ingenieure. (2015). *Systematische berechnunghochbeanspruchter schraubenverbindungenzylindrische einschraubenverbindungen* (techreport No. 1). Beuth Verlag GmbH.

A Appendix

A.1 Head Friction Model

```
model HeadFriction
  extends
      PartialEl..TwoFlangesAndSupport2;
  parameter Real mu_pos[:, 2]=[0, 0.5];
  parameter Real peak(final min=1) = 1;
  parameter Real dW(final min=0);
  parameter Real Dki(final min=0);
  parameter SI.Force fn_max(final min
      =0, start=1);
  extends
      Rot.Interfaces.PartialFriction;
protected
  parameter Real frad(final min=0) = (
      dW+Dki)/4;
equation
  mu0 = interpolate(mu_pos[:,1], mu_pos
      [:,2], 0, 1);
  w_event = w_relfric > 0;
  phi = flange_a.phi - phi_support;
  flange_b.phi = flange_a.phi;
  w = der(phi);
  a = der(w);
  w_relfric = w;
  a_relfric = a;
  flange_a.tau+flange_b.tau-tau = 0;
  fn = fn_max*f_normalized;
  tau0 = mu0*frad*fn;
  tau0_max = peak*tau0;
  free = fn <= 0;
  tau = if locked then sa*unitTorque
      else if free then 0 else frad*fn*(
    if startForward then
      interpolate(mu_pos[:,1], mu_pos
          [:,2], w, 1)
    else if startBackward then
      (-interpolate(mu_pos[:,1], mu_pos
          [:,2], -w, 1))
    else if pre(mode) == Forward then
      interpolate(mu_pos[:,1], mu_pos
          [:,2], w, 1)
    else (-interpolate(mu_pos[:,1],
        mu_pos[:,2], -w, 1)));
end HeadFriction
```

A.2 Embedment Model

```
model Embedment
  ...
  parameter SI.Distance L_max = 0.01;
  parameter SI.Time tauT = 0.5;
  parameter SI.Force fmax = 10;
  parameter SI.Force fmin = 10;
equation
  flange_a.s = s - L / 2;
```

```
  flange_b.s = s + L / 2;
  0 = flange_a.f + flange_b.f;
  der(L) = if flange_b.f < -fmin then
      -1 / tauT * (L - L_max * (
      flange_b.f / fmax)) else 0;
end Embedment;
```

A.3 Thread Model

```
model IdealThread
  extends PartialElementaryRotational..
  ToTranslational;
  parameter Real pitch(final unit="m",
      start=1);
equation
  (flangeR.phi - internalSupportR.phi)*
      pitch/(2*pi) = (flangeT.s -
      internalSupportT.s);
  0=flangeR.tau+flangeT.f*pitch/(2*pi);
end IdealThread;
```

A.4 Three Phase Bolt

```
model ThreePhaseBolt
  extends ..Interfaces..
  PartialCompliantWithRelativeStates;
  parameter
      SI.TranslationalSpringConstant
      c_lin(final min = 0, start = 1);
  parameter
      SI.TranslationalDampingConstant d(
      final min = 0, start = 1);
  parameter SI.Position s_rel0 = 0;
  parameter SI.Position s_rel_phase2;
  ...
algorithm
  s_rel2:= s_rel0 - s_rel_phase2;
  c_qua:= c_lin/(2*(s_rel0-s_rel2));
  s_rel1:= s_rel2-((s_rel2-s_rel0)/2);
equation
  contact =s_rel < s_rel0;
  linear =s_rel < s_rel2;
  f_c_lin =-c_lin*abs(s_rel - s_rel1);
  f_c_qua =-c_qua*abs(s_rel-s_rel0)^2;
  f_c =smooth(1, noEvent(if contact
      then f_c2 else 0));
  f_c2 =smooth(1, noEvent(if linear
      then f_c_lin else f_c_qua));
  f_d2 =if contact then d*v_rel else 0;
  f_d =smooth(0, noEvent(if contact
      then (if f_d2 < f_c then f_c else
      if f_d2 > -f_c then -f_c else f_d2
      ) else 0));
  f =f_c + f_d;
  lossPower = f_d*v_rel;
end ThreePhaseBolt;
```

Numerical Investigation on Performance of Gas Turbine Blade Effects of Simulation Models and Blade Geometry

Heng Hu, Narmin Hushmandi, Magnus Genrup

Energy Science department, Lund University, Box 118, 22100 Lund, Sweden
heng.hu@energy.lth.se

Abstract

With a significant impact on turbomachinery blade performance, surface curvature distribution becomes one of the essential factors in the design of high-efficiency blades. This study focuses on applying computational fluid dynamics (CFD) to evaluate turbine rotor blade performance. The main aim is to analyze the influence of incidence and geometry shape on the performance of a gas-turbine blade in two dimensions. To achieve this, an investigation was conducted to identify a suitable turbulence model for this case, with two turbulence models combined with two different solvers explored in ANSYS Fluent: Realizable k-ε model in pressure and density based solver; k-ω shear stress transport (SST) model in pressure and density based solver. The blade total pressure loss across different blade exit Mach numbers is the comparison factor, with validation against experimental data. Subsequently, the chosen pressure-based k-ω SST model mode is used to study the performance of various air inflow incidence angles and compare two different blade geometries. In this paper, two geometries, Geometry 1 and Geometry 2, were designed by setting two different exit blade angles, β_2=79.5° and β_2=70° respectively, while the inlet blade angles have the same value, β_1=48.8°. Furthermore, the effect of varying air inflow incidence angles between -48.8° and 10° on the blade performance distribution is also investigated. Within the studied range, the inflow incidence angle of 10° is found to have the best performance in terms of turbine work output. On the other hand, the blade performance of Geometry 2 appears superior to Geometry 1.

1. Introduction

The blade geometric profile is designed to determine the efficient aerodynamic performance. Some principal aerodynamic objectives of a turbine blade design are: the blade angles at the inlet and exit must be correctly matched to the fluid flow angles; the throat area determines the flow capacity and must be sized correctly. Besides, Blade surfaces curvature and changes in curvature should be limited, consistent with the necessary turning of the flow in blade passage. So the blade design plays an essential role in the full design process. In turbomachinery, quality blade design is an integral element to efficient aerodynamics (Lebele-Alawa et al., 2008), which can affect the entire blade row's performance, affecting the overall machine efficiency (Fast M et al., 2009). In particular, blade curvature distribution has been shown to influence boundary-layer characteristics, determining blade losses and efficiency (Korakianitis et al., 1993).Even though the field of blade curvature is relatively mature, the potential benefits of sizeable industrial cost-savings and environmental impact from even a tiny efficiency improvement have been sustaining the keen interest in work in this area.

The turbine portion in gas turbine systems extracts work from the combusted gases to power the compressor stages and drive other loads. As such, considerable effort has been poured into turbine blade research to attain maximum extraction of the valuable work output. Additionally, to further push the upper bound of turbine efficiency, much research has been done on limiting flow separation that contributes to decreases in work output (Korakianitis T P. et al., 1989) Unsurprisingly, the surface curvature of the blade determines its loading distribution, forming a crucial factor in controlling flow separation (Nemnem et al., 2014). In further detail, a smooth curvature distribution at the blade leading edge has been found to prevent the formation of separation bubbles, thus suppressing the flow separation (Song Y et al., 2014).

This paper considers how blade parameterization affects profile losses and loading diagrams in a numerical approach, including blade angles, incidence, etc. It is worth highlighting that one main objective of the paper is to explore and compare different numerical models against a set of experimental data to identify a suitable model for this application. Designing a turbine blade geometry typically starts from a one-dimensional approach before moving on to two-dimensional (2D) and eventually developing into 3D in the final phase. This paper involves the 2D round, which is practical

regarding the scope and computational cost. Coupled with the computationally economical RANS models, the approach reflects a common means adopted in industries to offer a quick blade diagnostic tool for trend studies. The selected RANS model is subsequently applied to the turbine blade design to relate the modified parameters to the selected blade performance indicators.

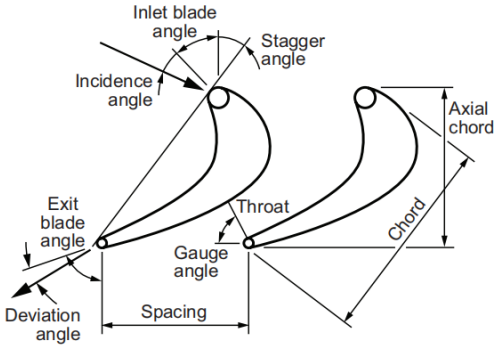
In addition, this paper also seeks to demonstrate the key role of CFD in both industry and academic research for turbine blade design. CFD today is an indispensable part of blade design as it provides a performance evaluation for a particular geometry. It can also be used to pick out the blade locations requiring modifications, forming an iterative part of the design process to optimize surface geometry and loadings.

2. Numerical Simulation

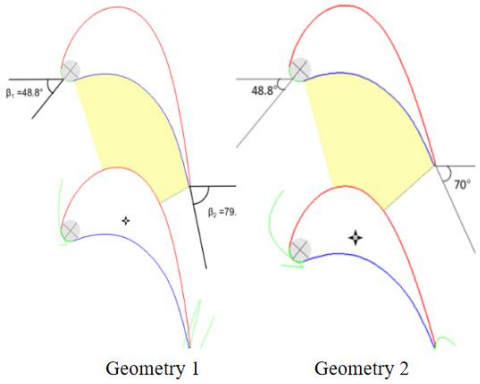
For this research, the fluid flow parameters, e.g., velocity magnitude, pressure, temperature, and Mach number, around the turbine blade will be simulated using numerical methods. Two turbine blade geometries with different exit blade angles were generated to examine the influence of blade geometry. The ANSYS ICEM CFD was used to mesh the geometry. The commercial CFD software, ANSYS Fluent, is used for solving and post-processing.

2.1. Turbine Blade Geometries

This paper aims to find the influence of different solvers, incidence angle, and shape of turbine blades. For simplification of simulation cost, 2D geometries are selected for this study. Figure 1(a) shows the schematic diagram of blade section parameters. AxCent of Concepts NREC design tools software provides a good way to generate the geometries by determining the blade section parameters (such as inlet/exit blade angles, stagger angles, gauge angles, wedge angles, chords, and pitch).



(a) Schematic diagram of blade section parameters



(b) Schematic of blade sections for two geometries.

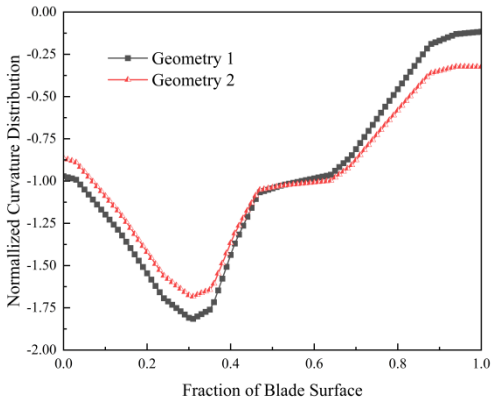
Figure 1: Schematic of the 2D blade.

In this research, two blades were designed by setting two different β2: exit blade angle, β2=79.5°, and β2=70° respectively, while the β1: inlet blade angle have the same value, β1=48.8°. The two blade geometries are shown in Figure 1(b). Geometry 1 is a reference blade profile of Atlas (Mee D J et al., 1992). Since Geometry 1 has experimental results, most of the simulations in this paper are based on Geometry 1. In order to study the influence of blade geometric parameters on blade performance, based on Geometry 1, modifications are done to the exit blade angle to obtain Geometry 2. The remaining important blade parameter values of Geometry 1 and Geometry 2 are shown in Table 1.

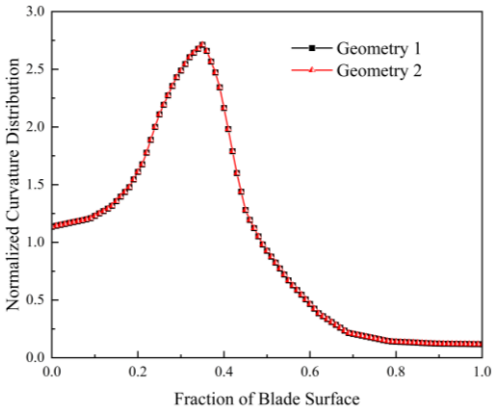
Table 1: Blade parameters.

Parameters	Geometry 1	Geometry 2
Chord (m)	0.0474	0.0474
Stagger Angle(°)	-37.8	-29.8
Pitch/Chord	0.7593	0.7597
Axial Chord (m)	0.03745	0.0411

The Normalized curvature distribution of the two geometries generated in AxCent showing in Figure 2, which shows that, in the two cases, the blade curvature is continuous and smooth. The blade curvature distribution of the pressure side for the two geometries differs, while the blade curvature distribution on the suction side coincides.



(a) Pressure Side



(b) Suction Side

Figure 2: Curvature distribution of the 2D blade sections for two geometries.

2.2. Mesh Generation and Boundary Conditions

The computational domain of fluid is divided into three parts: the upstream and downstream domains and the blade domain. The computation domain is extended upstream and downstream to achieve a fully developed flow. The inlet of the numerical domain is extended 0.8 times axial chord of the blade upstream of the blade leading edge, and the exit is extended 1.25 times the axial chord downstream of the blade trailing edge. Four types of boundary conditions are presented to solve a blade cascade: wall, periodic, inlet, and outlet (Moshizi S A et al., 2014). Using the periodic boundaries for the blade cascade is common in CFD to reduce the computational domain size and thus decrease the time and memory cost for processing (M. Mahmoudi et al., 2005), so just one blade passage simulated in all simulations. For the inlet and out boundary conditions, pressure-inlet and pressure-outlet are adopted, respectively. No-slip condition is used for the blade wall. The values of the boundary conditions applied to the cases are presented in Table 2.

Table 2: Boundary conditions.

Parameter	Value
Inlet/Outlet Total Temperature	1046(K)
Inlet Absolute Total Pressure	211325(Pa)
Outlet Absolute Static Pressure	126325(Pa)
Outlet Mach Number	0.7-1.1

With good quality of the structured mesh, using a higher-order discretization scheme, the solver solution would have a higher convergence rate and precision. Therefore, the software ICEM CFD was used to generate the structured grids for calculation domains around the blade surface. Figure 3 shows the meshes of Geometry 1 employed in the computational domain. The whole grid was structured with an O-H type of mesh, using the O-type mesh around the blade and H-type everywhere else. The meshes are refined for the near-wall treatment and are expected better to handle the complex turbulent flow around these areas and enhance computational accuracy. The first layer grid near-wall is 0.002 mm, and the y+ value is around 1.

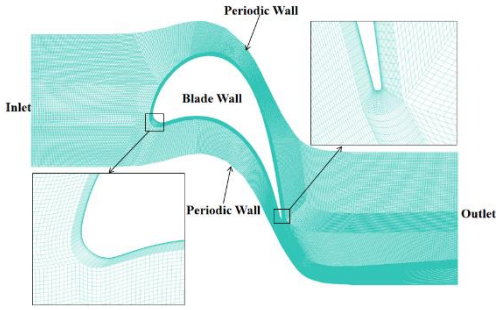


Figure 3: Numerical domain and mesh for the analysis of the geometry 1.

2.3. Grid Independence Study

A mesh independence analysis was done using various mesh densities to study the effect of grid resolution on the accuracy of numerical results. For this purpose, the grid resolution was increased until the blade total pressure loss had no significant variations.

This paper uses the total pressure loss coefficient to characterize the blade total profile loss. The definition of the total pressure loss coefficient:

$$Y_p = \frac{p_{01,in} - p_{02,out}}{p_{01,in} - p_{1s,in}} \quad (1)$$

Where $p_{01,in}$ represents the blade inlet total pressure, $p_{02,out}$ is the mass-weighted average total pressure at the blade section where extended 0.8 times the axial chord downstream of the blade trailing edge.

Nine grid sizes in the range of 5000 to 134 000 structured cells are evaluated. Figure 4 illustrates the variations of blade total pressure loss for the different grids. Due to the importance of computational efficiency, the mesh with 114 400 structured cells was chosen for Geometry 1; Geometry 2 has meshed with the same method, and the number of grids is comparable at 118 650 structured cells.

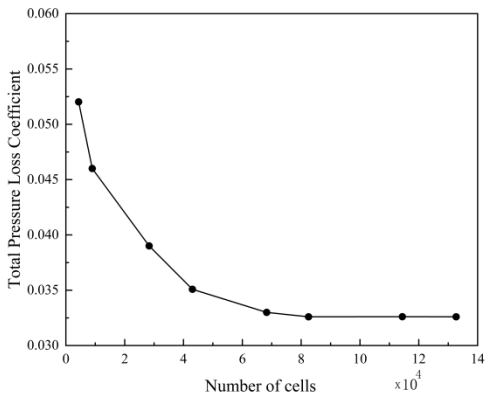


Figure 4: Mesh independence study of Geometry 1

2.4. Solver Settings

The governing equations for viscous compressible fluid are the continuity, Navier-Stokes momentum, and state equations. The equations are discretized with the finite volume method. Firstly, four different RANS models in Fluent - namely the Pressure-based k-epsilon method, Pressure-based SST k-ω method, Density-based Realizable k-ε method, Density-based SST k-ω method were used to solve the discretized equations. For the spatial discretization of pressure-based solver cases, the second-order scheme is used for pressure, and the second-order upwind scheme is used for momentum and turbulent kinetic energy terms. For the spatial discretization of density-based solver cases, the second-order upwind scheme is used for flow and turbulent kinetic energy terms. Besides, ideal gas model has been chosen because the physical fluid is a compressible fluid. Turbulence is chosen to be modeled using the Realizable k-ε and SST k-ω models due to the corresponding theoretical strengths in providing realistic results and superior performance for complex flows (like adverse pressure gradient and separated flows), respectively. The convergence of the solution is monitored by checking the residuals of the numerically solved governing equations, which use the absolute scale of residuals to converge until 1E-6.

In Fluent, the pressure-based solver is developed from the original separate solver, which solves the momentum, pressure correction, energy, and other scalar equations in sequence, such as the turbulence equation. Unlike before, the pressure-based solver also adds a coupling algorithm, which can be freely switched between the separation and coupling solutions. The coupling solution is to solve the aforementioned momentum and pressure correction simultaneously and then solve energy, component equations, and other scalar equations, such as Turbulence equations, etc., which have fast convergence speed but require more memory and calculation.

The difference between Pressure-based and Density-based: First, the pressure-based solver was mainly used for the solution of low-speed incompressible flow, while the density-based method was mainly designed for high-speed compressible flow, but now both ways have been extended to solve a large flow velocity range method. Second, the density-based solver was developed from the originally coupled solver. It simultaneously solves the continuity, momentum, energy, and component equation, then solves the turbulence and scalar equations. As a result, the density-based solver has a fast convergence speed and requires more memory and calculation time than the pressure-based solver!

As a result, the density-based solver is expected to take longer computational time per iteration. According to observation, for the studied Geometry 1, the density-based simulations took significantly longer to run and converge. The coupled algorithm is used for the pressure-based solver because of its higher accuracy. As the simulation is 2D and the grid resolution is not large, the trade-off in computational time is insignificant.

2.5. Model Validation

The experimental results of Atlas, involving the same blade design as Geometry 1, were used to validate the numerical solutions. The blade profile loss against different Mach numbers was profiled in the actual experiment. The inlet airflow direction specification method was set to be normal to the boundary, corresponding to an incidence angle of -48.8°. Additionally, the specification method for both inlet and outlet boundary conditions was based on turbulence intensity and turbulence viscosity ratios of 5% and 10, respectively. Different exit Mach numbers were obtained by adjusting the outlet pressure value.

Figure 5 shows the numerical blade losses across various exit Mach numbers found via the four model-solving cases close to experimental data. The models are pressure-based SST k-ω, density-based SST k-ω, pressure-based Realizable k-ε, and density-based Realizable k-ε. Although the numerical data do not overlap entirely with the experimental ones, the differences are within a reasonable range. The changing trend of the experimental and simulated structures is basically the same and have some slight difference. The plotted loss coefficients are for exit Mach numbers between 0.7 to 1.1. It is seen from the plot that, up to a specific exit Mach number (Ma=0.9), the total pressure losses are low. However, after Ma=0.9, the total pressure loss increases rapidly. The sudden increase is due to the appearance of shocks inside the blade channel. As the exit Mach number continues to grow to about 1.05, the simulation results show that the total pressure loss increases slowly with the exit Mach number, while the experimental results still increase significantly with the increase of the exit Mach number.

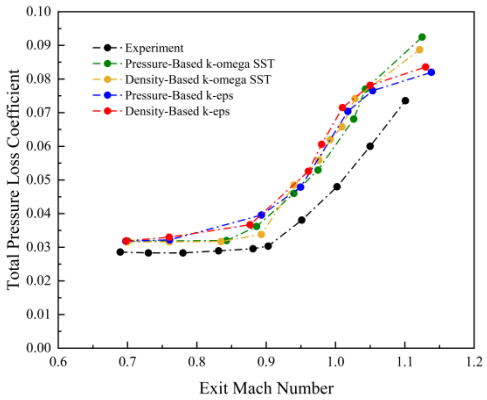


Figure 5: Comparison between numerical and experimental blade profile loss for Geometry 1.

The differences may have come from the following reasons. Firstly, the compressibility of air and the capture of shock waves in the simulation calculations may still differ from the actual problem; some factors, like the turbulent vortex in 3D space and dissipation in the third direction, cannot be captured by the 2D simulations in this research. Secondly, applying constant static pressure at the outlet affects the total pressure loss values. Besides, the empirical data measurement position and statistical approach may differ from the data collection of our numerical approach. Finally, this study is mainly on a transonic turbine blade, with exit Mach number ranging from about 0.7 to 1.1, for which it is notoriously difficult to get a precise solution with RANS models.

Similarly, the differences between the four cases are also slight. The data sets for density-based and pressure-based SST k-ω models are close to each other, while the density-based and pressure-based Realizable k-ε models also have a similar phenomenon. According to the literature (Corriveau D et al., 2007), the profile loss will vary slowly for high Mach numbers, which also can be observed in our current numerical results. When the Mach number is above almost 1.05, the density-based and pressure-based Realizable k-ε model has less variation than the SST k-ω models. So in terms of growth trend, the SST k-ω model is closer to the experimental growth trend than the Realizable k-ε model.

3. Results and Discussions

3.1. Comparison across Models

Before further research on this turbine blade, some variables, like Mach number, pressure, obtained in the model validation section were identified as key parameters. In addition, comparing the differences caused by different model-solver cases is also critical. Using geometry 1, we have done some simulations to explore the influence of model and solver selection. This study analyzes the conditions of different outlet Mach numbers 0.76 and 1.1, as shown in Figures 6-11. The most apparent difference between these two cases is the observation of shock waves. When the outlet Mach number is 1.1, the shock wave is obvious in all four models, as seen in Figure 6 and 7, while the flow field of four models is subsonic when the outlet Mach number is 0.76, as shown in Figures 9 and 10.

Figure 6 shows the contour of the exit Mach number of 1.1 under the same turbulence models with different solvers. Specifically, the density-based and pressure-based Realizable k-ε models have similar Mach number distributions, including the value range and position of the shock wave, while the density-based and pressure-based SST k-ω models capture a more severe shock. As it is seen from zoom-in figures, near the suction side of the outlet blade wall under the same solver, the SST k-ω model shows a better distribution of the reflected oblique shocks and expansion waves near the blade wall. Theoretically, the Realizable k-ε model might face inaccuracies for complex wall-bounded flows, such as predicting the early onset of flow separation or the inability to accurately capture turbulent shock wave-boundary layer interaction. Thus SST k-ω might be more reliable in this respect. The Mach number around the trailing edge of the turbine blade is small, which means the velocity is small, resulting from some adverse flow decrease in kinetic energy.

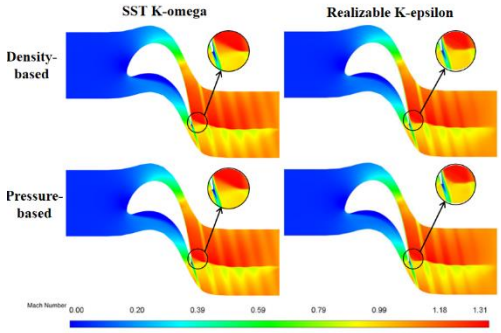


Figure 6: The Contour of Mach number of four models with exit Mach number 1.1.

Figure 7 and 8 show the contour of static temperature and total pressure of four models, respectively. Besides, in both figures, the 'inclined strip' near the outlet part is caused by the blade wakes, and the 'strips' are caused by other turbine blades present with the periodic boundary condition selection.

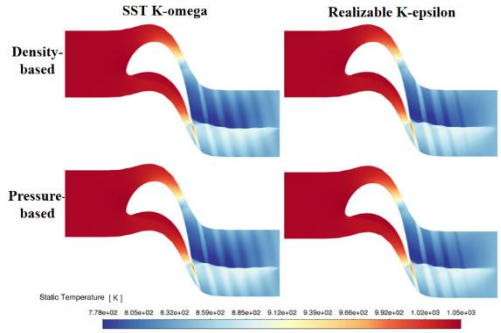


Figure 7: The static temperature of four models with exit Mach number 1.1.

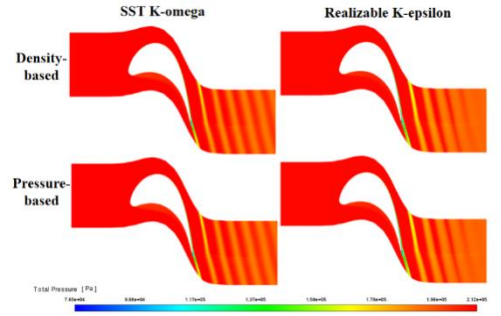


Figure 8: The Total Pressure of four models with exit Mach number 1.1.

As the outlet Mach number was decreased to 0.76, the entire flow field becomes subsonic, as shown in Figures 9. The maximum Mach number changed from about 1.31 to 0.85. Besides, the Mach number in the blade's trailing edge is more uniform, and there is no shock wave in the field.

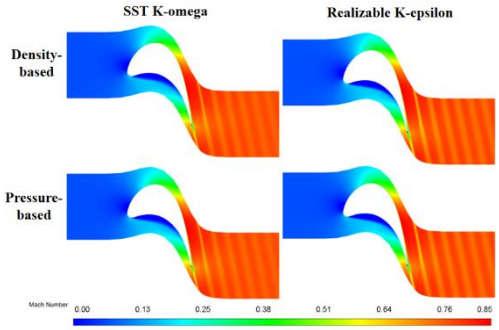


Figure 9: The Contour of Mach number of four models with exit Mach number 0.76.

A low Mach number means the kinetic energy loss is slight. Therefore, the static temperature in the trailing edge of the blade increased in a smaller area, and the distribution near the outlet is more even, as shown in Figure 10. The average temperature is higher than the case with a high Mach number. This is because the two cases with the same inlet temperature have higher Mach number with shock waves increasing more loss so that the temperature will be higher.

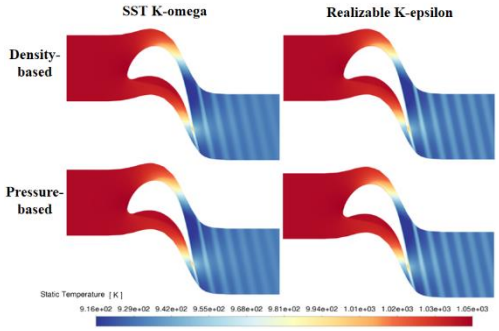


Figure 10: The static temperature of four models with exit Mach number 0.76.

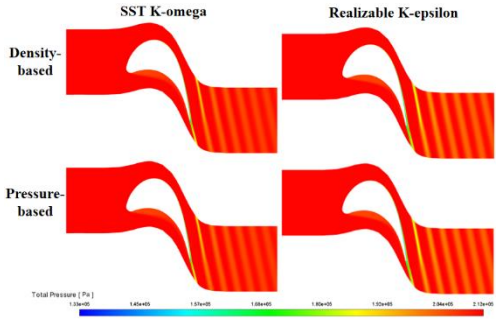


Figure 11: The Total Pressure of four models with exit Mach number 0.76.

The temperature data (Figures 7, 10) also offers information pertinent to the design of turbine blades. In a gas turbine, the turbines would endure some of the harshest operating temperatures. On top of revealing the temperatures the blades will be subject to, the temperature distribution around the blade can potentially allow the designer to measure how design changes might lead to temperature changes in the flow field and uneven distributions, if any. Furthermore, additional design measures such as cooling channels and thermal barrier coating can be incorporated to target the areas with peak temperatures.

Figure 11 shows the total pressure of four models when the outlet Mach number is 0.76. It can be seen that the total pressure drop in Figure 11 is significantly smaller than that in Figure 8. Therefore, the shock wave is one of the significant sources of loss. From this, it can be concluded that the blade curvature distribution should be in a way to minimize the effect of losses with having oblique and expansion waves in the flow.

So far, the results of the density-based and the pressure-based are not much different, and the pressure-based costs less computational time. Besides, turbulence models, SST k-ω model compared to Realizable k-ε model has more ability to accurately capture turbulent shock wave-boundary layer interaction. Therefore, the pressure-based SST k-ω model is thought to be more realistic for this study. Therefore, the computations performed and the results presented below are performed entirely with pressure based SST k-ω model.

3.2. Influence of Incidence

To study the influence of incidence on the flow around the turbine blade, different incidence angles are applied at the inlet with the same geometry 1, setting the same inlet and outlet pressure, which exit Mach number is almost 0.88, and the maximum Mach number is around 1, as shown in Figure 12. Increasing the incidence angle will bring about the translation of the low-speed stagnation regions from the blade leading edge to the pressure side and increase in inlet Mach numbers.

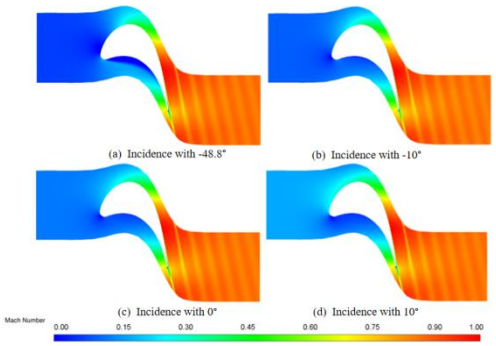
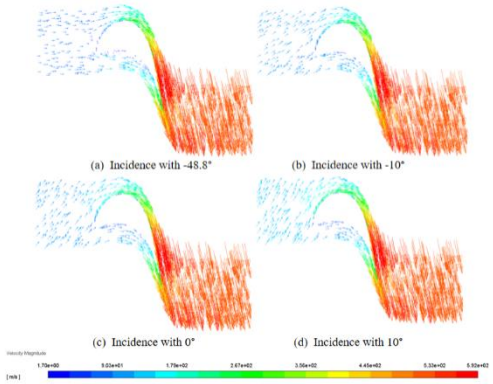
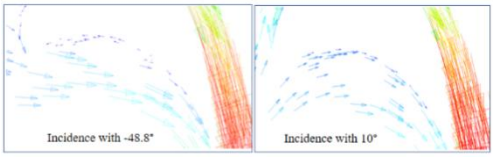


Figure 12: Contour of influence of incidence on Mach number in Geometry 1



(a) Contours of velocity vectors

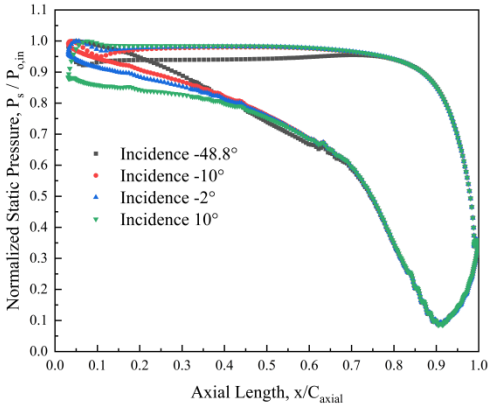


(b) Zoom-in Figures of two incidence angle

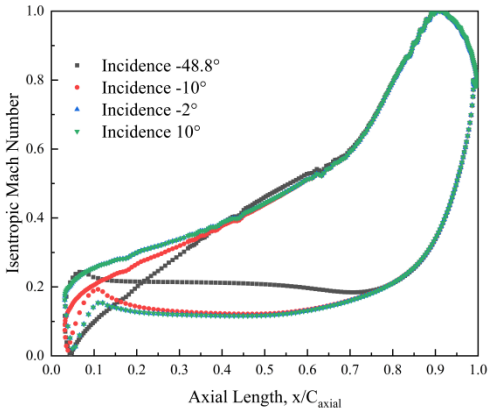
Figure 13: Contours of velocity vectors showing effect of incidence angle on the Geometry 1.

Figure 13 shows the contour of the velocity vector and zoom-in figures. Figure 13(a) compares the contour of different incidence angles on Mach number in Geometry 1. From the zoom-in figure (b), showing some specific details, there is a backflow and vortex near the pressure side boundary region when the incidence angle is -48.8º, which may cause more losses. However, as the incidence angle change to 10º, there is no reverse flow and vortex.

Isentropic Mach number and static pressure values along the blade surface normalized with total inlet pressure p_s⁄p_(0,in) are plotted against the axial distance from the leading edge of the blades (x⁄C_axial), where C_axial is the blade axial chord, is adopted to better evaluate performance due to the small numerical differences for blade profile losses, which shows in Figure 14. The small 'loop' in Figure 14 (a) when the incidence is -48º indicates negative work done by the blade from the flow before the position of 0.25. The net area enclosed by the curve can be related to the work output by the blade so that a larger one will indicate better work output. Besides, we can also know that over-expansion between the throat and trailing edge exits some diffusion, with an increase in loss. Figure 14 (b) compares the influence of different incidences on isentropic Mach number distribution. It can be seen from the Figure 14 (b) that the maximum isentropic Mach number of the four cases does not exceed 1. The blade's performance depends largely on the amount of diffusion and diffusion rate, which is also important when we further optimize the blade performance in the future.



(a) Normalized Static Pressure



(b) Isentropic Mach Number

Figure 14: Influence of incidence on blade loading along the surface of the blade in Geometry 1.

Identifiable from Figure 14, the incidence of 10º gives the largest work output (within the studied range). This information could be incorporated either into the design of the trailing edge blade angle of the stator stage upstream of the concerned rotors or the orientation of the rotor blades to obtain the desired incidence angle.

On the other hand, gas turbines frequently operate at off-design conditions. The airflow entering each turbine stage can be far from the designed incidences at off-design conditions. A transonic or low supersonic flow coupled with a large incidence, possibly leading to a significant flow separation on the turbine blade, would pose a real challenge for turbine designers. The data obtained from incidence angle simulations can be useful for performance analysis of off-design operating conditions.

3.3. Effects of Exit Blade Angle and Incidence

Figure 15 shows the Mach number distribution of the blades in Geometries 1 and 2.

The performances of Geometry 1 and Geometry 2 with the incidence of 10º and -10º were compared to observe the influence of the modified exit blade angle on the blade performance. The results are shown in Figures 16.

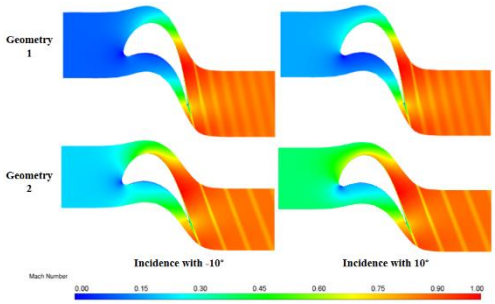
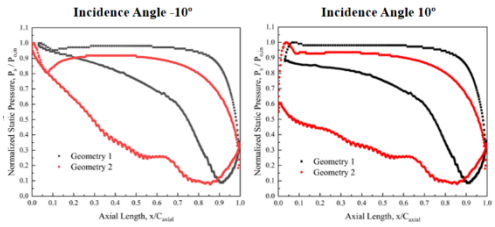
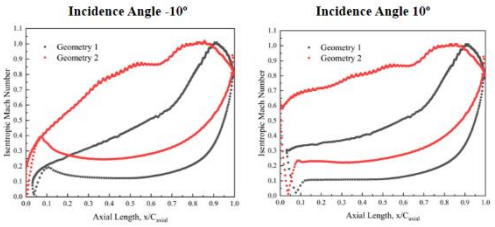


Figure 15: The Mach number distribution of the blades in Geometries 1 and 2.



(a) Normalized Static Pressure



(b) Isentropic Mach Number

Figure 16: Influence of incidence on blade loading along the surface of the blade in Geometry 1 and 2.

Figure 16(a) shows the normalized static pressure distribution along the blade surface. The enclosed curve area of Geometry 2 for all studied incidence angles is bigger than those of Geometry 1.

From Figure 16(b), Geometry 2 has a smaller maximum isentropic Mach number for the same incidence angle than Geometry 1. Besides, the position of peak velocity on the suction side is earlier, so the diffusion rate is less, with a decrease in loss.

By this measurement, the performance of Geometry 2 is better than Geometry 1, which indicates that decreasing the exit blade angle has led to improving the performance of the current turbine blade.

4. Conclusion

This paper mainly focuses on applying CFD with a suitable turbulence model to evaluate turbine rotor blade performance. The results have shown the adequacy of the four RANS models – pressure and density-based Realizable k-ε and SST k-ω, in simulating the flow field trends for the Geometry blade design to a reasonable accuracy. The pressure-based SST k-ω model has been eventually picked as the model of choice due to the slightly better matching of the experimental data trends, capability to capture shock waves in the performed simulations more accurately, lower computational cost of the pressure-based solver, as well as SST k-ω model's theoretical superior ability to handle complex flows including those around turbine blades. So this paper finds a cost-effective CFD model that can predict performance trends with reasonable accuracy. It provides a more convenient and reliable method for performance evaluation of 2D turbine blade geometries.

Besides, different incidence angles are studied to see the influence on the blade performance. Within the study range, the inflow incidence angle of 10° is found to have the best performance in terms of turbine work output. Two geometries were designed by setting two different exit blade angles to observe the influence of the modified exit blade angle on the blade performance. The blade performance of Geometry 2 appears superior to Geometry 1.

Finally, Geometry 2 with an exit blade angle of 70º, coupled with the incidence angle of 10º (among the cases of incidence studied), has been shown to give the largest work output and fewer losses.

References

Lebele-Alawa B T, Hart H I, Ogaji S O T, et al. Rotor-blades' profile influence on a gas-turbine's compressor effectiveness[J]. Applied Energy, 2008, 85(6): 494-505.

Fast M, Assadi M, De S. Development and multi-utility of an ANN model for an industrial gas turbine[J]. Applied Energy, 2009, 86(1): 9-17.

Korakianitis T. Hierarchical development of three direct-design methods for two-dimensional axial-turbomachinery cascades[J]. 1993.

Korakianitis T, Papagiannidis P. Surface-curvature-distribution effects on turbine-cascade performance[C]//Turbo Expo: Power for Land, Sea, and Air. American Society of Mechanical Engineers, 1992, 78934: V001T01A044.

Korakianitis T P. Design of airfoils and cascades of airfoils[J]. AIAA journal, 1989, 27(4): 455-461.

Nemnem A F, Turner M G, Siddappaji K, et al. A smooth curvature-defined meanline section option for a general turbomachinery geometry generator[C]//Turbo Expo: Power for Land, Sea, and Air. American Society of Mechanical Engineers, 2014, 45615: V02BT39A026.

Song Y, Gu C W, Xiao Y B. Numerical and theoretical investigations concerning the continuous-surface-curvature effect in compressor blades[J]. Energies, 2014, 7(12): 8150-8177.

Mee D J, Baines N C, Oldfield M L G, et al. An examination of the contributions to loss on a transonic turbine blade in cascade[J]. 1992.

Moshizi S A, Madadi A, Kermani M J. Comparison of inviscid and viscous transonic flow field in VKI gas turbine blade cascade[J]. Alexandria Engineering Journal, 2014, 53(2): 275-280.

M. Mahmoudi, M. Ansari, Numerical investigation of turbine blade trailing edge flow ejection effects on mach number distribution of gas turbine blade surface. Using rng.k-e turbulence model[J].2005.

Corriveau D, Sjolander S A. Influence of loading distribution on the off-design performance of high-pressure turbine blades[J]. 2007. JNL/27.2.97.

Insight into the thermodynamic model for reforming of methane over nickel catalyst

Rakhi [a,*], Binod R. Giri [a], Vivien Günther [b], Fabian Mauss [a]

[a] *Brandenburg University of Technology, 03046 Cottbus, Germany,*
[b] LOGE AB, 03044 Cottbus, Germany
[*] *rakhi.rakhi@b-tu.de*

Abstract

The reforming of light hydrocarbons to produce synthesis gas, H_2 and CO, is an important intermediate for manufacturing valuable basic chemicals and synthesis fuels. In order to understand these reforming processes better, elementary step reaction mechanisms are developed. In the available literature, the surface reaction mechanisms are usually achieved with the help of reaction kinetic parameters without using the thermochemistry of the species referred to kinetic models due to the unavailability of the thermochemistry of the intermediate species involved in the multi-step reaction mechanism. In this work, investigations are made to obtain the thermochemistry of the intermediate species to establish thermodynamic equilibrium in order to develop a thermodynamic model for steam reforming of methane over nickel. The thermochemistry of the surface bound species is taken from different sources available in the literature and after that a detailed sensitivity analysis is performed to match the results with experiments. The simulation set up is adapted from the literature experiments given in [1]. The results produced with the one-dimensional tool using the thermodynamic model developed in the present investigation consisting of 21 reversible reactions are compared with the kinetic scheme with 42 irreversible reactions from reference simulation along with their experimental results. Both the models show some major differences in the reaction pathways which provides a useful insight into the key rate determining steps and needs further investigations.

Keywords: Nickel-based catalyst; One-dimensional modeling; Thermodynamic model; Steam reforming;

1. Introduction

A crucial chemical process providing synthesis gas (H_2 and CO) is steam reforming of hydrocarbons [2, 3]. The reforming of steam plays an important role as a feedstock in many catalytic processes [1], such as, synthesis of methanol, oxo-synthesis, and Fischer-Tropsch synthesis. The methane or gas (natural) reforming is the most prominent and widely used industrial steam reforming process [1, 3, 4].

There are several kinetic models available in literature to study reforming processes. Quiceno et al. [5] described the steam reforming of methane accompanied by water-gas shift reactions on a $Ni/MgAl_2O_4$ catalyst by intrinsic rate equations derived from a Langmuir-Hinshelwood mechanism. A catalytic sequence for reactions of CH_4 with CO_2 and H_2O on Ni/MgO catalysts is considered by Wei et al. [6]. A microkinetic model for steam reforming reactions over a $Ni/MgAl_2O_4$ catalyst is investigated by Aparicio et al. [7] by reactions for CO_2 reforming of methane and deactivation by carbon formation.

Other investigations focusing on reduction of the cost of synthesis gas production are partial oxidation over noble metal catalysts [8–11] as well as CO_2 reforming [12–14] of natural gas to synthesis gas. Several investigations considered the sequence and interaction of the reaction routes to understand the reaction mechanism of synthesis gas formation from methane. A direct catalytic partial oxidation route has been studies by Hickman et al. [11] and the indirect route in [8–10, 15]. The steps for steam reforming for the catalytic partial oxidation of methane over platinum and rhodium are published in [11, 15–17]. In [18, 19], the reaction kinetics of methane steam reforming over nickel catalyst has been extensively investigated experimentally and theoretically.

All the investigations in literature consider a kinetic model with all the direct/irreversible reactions. The detailed surface reaction mechanism in these models requires specification of Arrhenius parameters for all the forward reactions involved in the mechanism. In these models, even the reverse rates are calculated with the help of kinetic parameters without using the thermochemistry of the intermediate species involved in the mechanism. Hence the equilibrium can be achieved by modifying the Arrhenius parameters. The prime reason to follow this approach is because of the unavailability of the thermodata of the surface species.

However, the optimum approach to establish thermodynamic equilibrium would be to use the thermodata of the surface species to calculate the reverse rate expressions. The Arrhenius parameters are more accurate to define the speed of the reactions whereas the thermochemistry is more important to achieve the equilibrium. This motivated us to develop a thermodynamic model where the kinetic parameters are given only for the forward reactions and then the rate

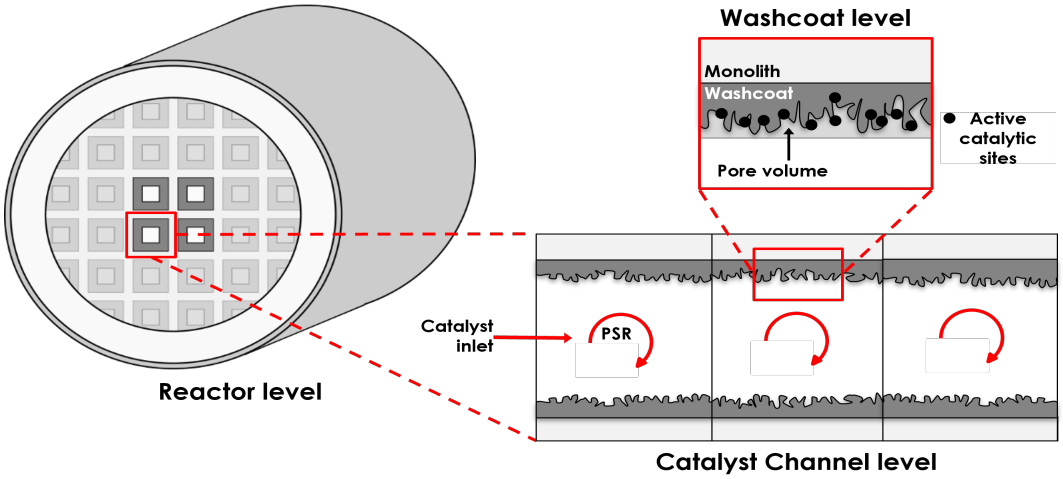


Figure 1: Schematic illustration of the one-dimensional modeling approach

expressions for the reverse reactions can be calculated by using the thermodata of the intermediate species.
The thermochemistry of the intermediate species available from this study can be further directly used in other reforming processes, for instance, dry reforming or oxidative reforming of methane over a nickel catalyst in order to understand and mitigate the coke formation problems encountered in these reforming processes at different conditions.

2. Simulation Set-up

The simulation set-up is taken from Maier at al. [1]. This allows us to re-calculate the results for the kinetic model with the LOGEcat for validation of the model as well as the surface reaction mechanism and later checking the predictive capability for the thermodynamic model. As done in the reference paper [1], the simulations are performed for different reactor conditions in terms of parameter as temperature while keeping the fuel ratio, flow rate and pressure same as [1]. The simulations are carried out at four temperatures, i.e., 920, 1020, 1120, and 1220 K.
The one-dimensional model, LOGEcat [20] is used for the simulations and the model is based on the single-channel 1D catalyst model applicable to the simulations of all standard after-treatment catalytic processes of combustion exhaust gas, for example, three-way catalyst (TWC), diesel oxidation catalysts (DOC), NO_x storage and reduction (NSR) catalysts and selective catalytic reduction (SCR) catalysts.
The single channel divided into a finite number of cells with Δx as their length is shown in Figure 1. As shown in the figure, each cell is treated as a perfectly stirred reactor (PSR). The pressure gradient along with inhomogeneity of the mixture can be neglected because the diameter of the catalytic channel is small. The external diffusion is modeled by a thin layer represented by a separate pore gas zone close to the wall. This pore layer is depicted by the area between the bulk gas and the washcoat as shown in Figure 1. For more details about conservation equations used in the model and the derivations of these equations, we refer the reader to [20, 21] and to understand the modeling approach, to our previous work [22–25].
A single channel being 1.0×10^{-2} m in length with a catalyst radius of 7.5×10^{-3} m which is uniformly divided into 25 cells is considered for the simulations with one layer of washcoat. The overall heat transfer efficiency factor, mass transfer efficiency factor and the efficiency factors for the surface chemistry are taken as one. The surface site density for nickel is 2.6×10^{-5} mol/m^2 [3]. The surface area per catalyst length is selected 6.9×10^{-3} m^2/m by performing the sensitivity analysis to get the results comparable to the reference simulations and experiments. Argon dilution is used similar to [1].
The reaction mechanism for the simulations is taken from Maier et al. [1] which contains 6 gas-phase and 13 surface species in total along with 42 forward reactions referred as kinetic model. After validating the results produces with the LOGEcat using the kinetic model, a thermodynamic model is developed. In this model, only 21 forward reactions are used with the kinetic parameters taken from Maier et al. [1]. The kinetic expressions for the remaining 21 backward reactions are then calculated with the help of the thermochemistry for the intermediate species involved in the surface reaction mechanism. Since the thermochemistry of the surface bound species is not easily available, we have taken the thermodata from different sources [1, 26]. The thermodata from Maier et al. [1] is referred as DETCHEM and from Liu et al. [26] as RMG in upcoming sections.
The heat capacity, entropy and enthalpy for all the intermediate species involved in the surface mechanism have been summarised in [22] in Table 1 for both the sources, DETCHEM as well as RMG. We note that the enthalpy of formation of species vary in a wide range for the two sources, specially for the species, C(s), CH_3(s), CH(s), and CH_4(s). This hints towards the requirement of a further investigation to find the sensitive limits for thermochemistry of all the species. Nevertheless, the most sensitive reactions for dry reforming of methane using a kinetic model and the thermodynamic model have been given in [25], Table 3.

3. Results

The above mentioned 1D model is used to perform the simulations for various temperatures for the steam reforming of methane over a nickel catalyst. The kinetic parameters for the considered 21 forward reactions and the thermochemistry of the 13 surface species are taken from [1].

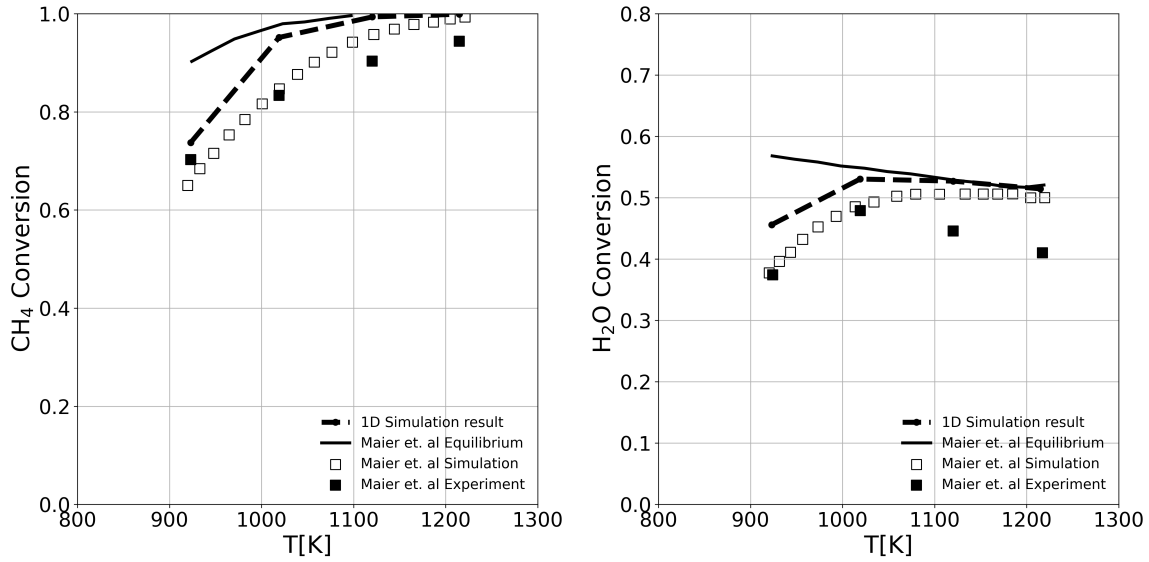


Figure 2: Methane and water conversion as a function of temperature along with the reference data. The 1D simulation results are shown for thermodynamic data from DETCHEM by replacing the species CO(s) from RMG data.

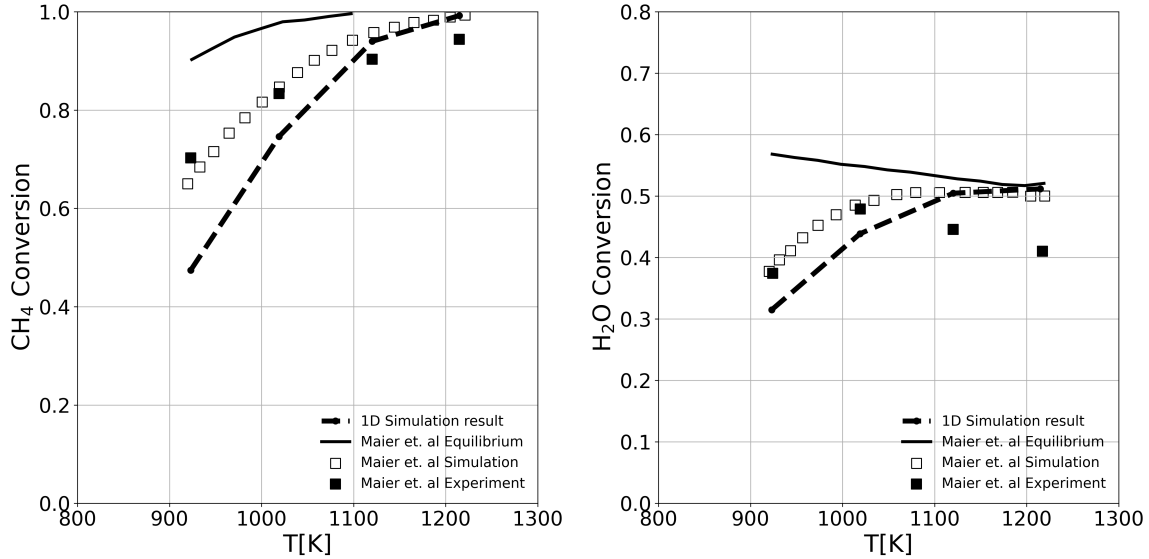


Figure 3: Methane and water conversion as a function of temperature along with the reference data. The 1D simulation results are shown for thermodynamic data from DETCHEM by replacing the species $CO_2(s)$ from RMG data.

As an initial check, we used the thermodata only from DETCHEM and check the methane and water conversion in the considered temperature range. The species concentration showed deviations (under predicted) and this motivated us to take the thermodata from RMG for all the intermediate species where the deviation in the conversion profile was noted more than the previous data set (still under predicted). So, in order to understand the impact of thermodata of the individual species, we next considered the thermodata for all the 13 species from DETCHEM and then replaced the thermodata of species one by one with RMG.

We note that the target for the thermodynamic model, presented in this paper, is to perform close to the reference experiments. However, there are several assumptions to develop the reference kinetic model [1] which is used as the base to develop the thermodynamic model. So, the results for our model are expected to remain close to the reference simulations.

3.1. Influence of CO(s)

The influence of only some of the site species are shown and discussed in this paper. Figure 2 shows the methane and water conversion as a function of temperature using the thermodynamic model along with the reference experiments and simulations with kinetic model from [1]. The thermodynamic model used the thermodynamic data for all the species from DETCHEM [1] and only for the species CO(s), the thermodynamic data is used from RMG [26]. The conversion for both the reactants is over predicted at all the temperatures.

For methane conversion in the range 900-1100 K, the over prediction is significant whereas, for higher temperature (1200 K) thermodynamic equilibrium is achieved. In case

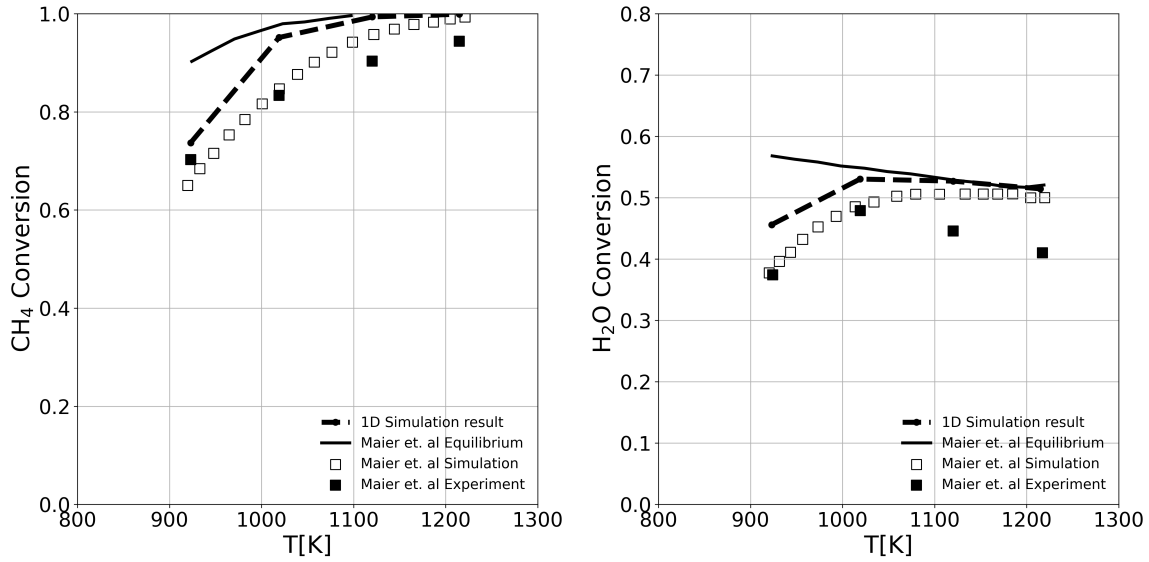


Figure 4: Methane and water conversion as a function of temperature along with the reference data. The 1D simulation results are shown for thermodynamic data from DETCHEM by replacing the species CO(s) and $CO_2(s)$ from RMG data.

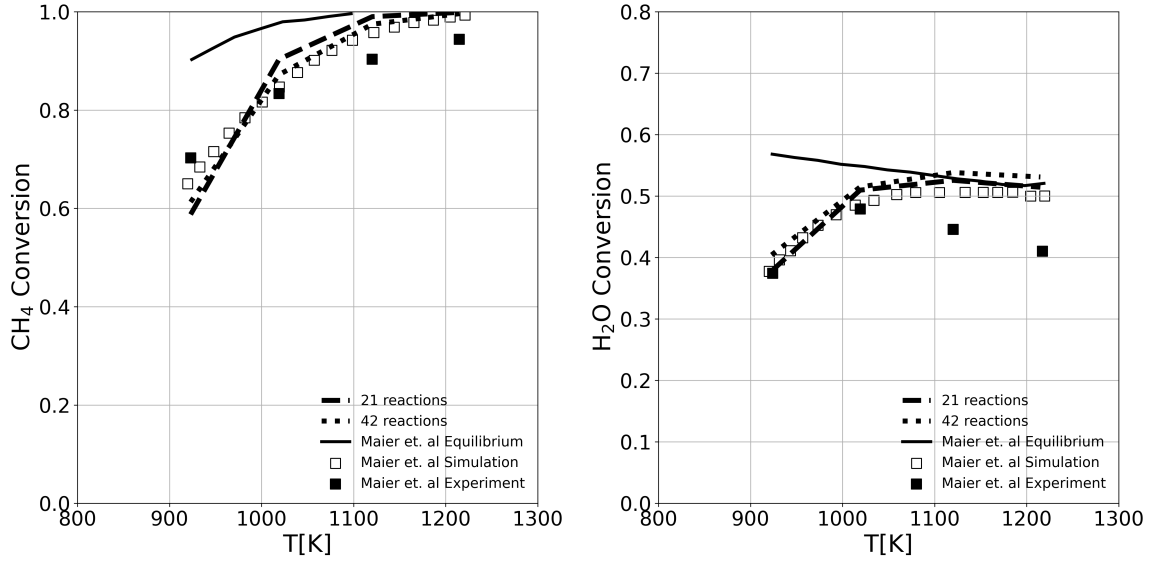


Figure 5: Methane and water conversion as a function of temperature along with the reference data with the kinetic model. All other parameters are kept fixed. 1D simulation results produced with the kinetic model (42 reactions) as well as thermodynamic model (21 reactions) are shown.

of water, the conversion is deviating from the reference simulations at low temperatures and shows agreement at temperatures above 1000 K. The water conversion computed with the 1D model as well as for reference simulations is away from reference experiments, specially at high temperatures.

3.2. Influence of $CO_2(s)$

The influence of the species $CO_2(s)$ shown in Figure 3 is completely different as compared with the CO(s) species. For both the species, methane conversion is in equilibrium at higher temperatures. The conversion of both the reactant species is under predicted at low temperatures by utilizing the thermodata for all the species from DETCHEM [1] and for $CO_2(s)$ species from RMG [26]. However, if the thermodata for both the species, CO(s) and $CO_2(s)$, is taken from RMG [26], shown in Figure 4, while all other species thermodata is taken from DETCHEM [1], then the species conversion is similar to the one explained above in Figure 2.

Such investigation was further performed for all the 13 surface species and we found that out of all the involved site species, CO(s) and $CH_4(s)$ are the most sensitive and play an important role to achieve conversion of the species comparable to the reference experiments. However, a direct replacement of the RMG species with DETCHEM could not compute the conversion correctly. This lead us to perform the enthalpy sensitivity analysis of the most sensitive species. The enthalpy of formation of the species CO(s) (thermodata for this species is taken from RMG) was further increased from 1 to 100 kJ to achieve the reactant conversion comparable with the reference experiments and simulations.

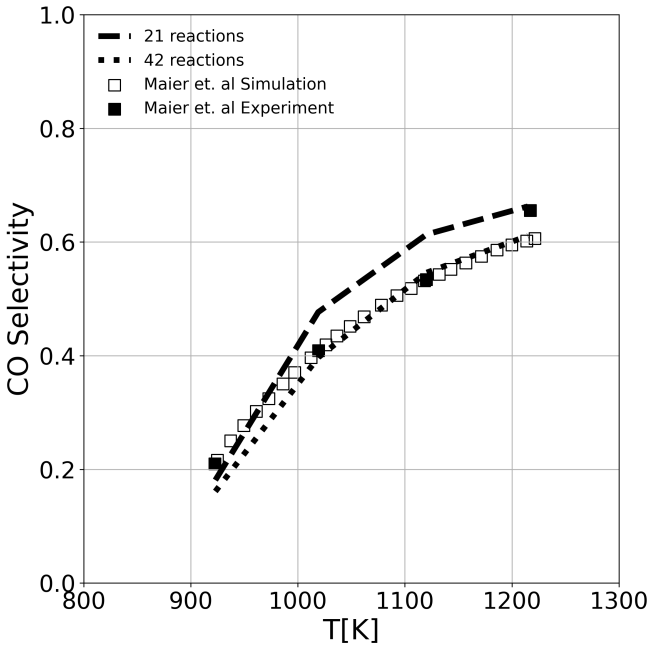


Figure 6: CO selectivity as a function of temperature along with the reference data. The reference data is using the kinetic model while the 1D simulation results are shown for both, the kinetic model and the thermodynamic model.

3.3. Comparison between kinetic and thermodynamic model

Further, the simulations are performed using LOGEcat with the kinetic model taken from Maier at al. [1] and with the thermodynamic model developed in the considered temperature range. Methane and water conversion as a function of temperature along with the reference experiments and simulations with the kinetic model is shown in Figure 5 keeping all other parameters fixed. In this figure, the LOGEcat results are shown for kinetic model using 42 reaction as well as for thermodynamic model using 21 reactions. 1D simulation results produced with the kinetic model (42 reactions) as well as thermodynamic model (21 reactions) are in good agreement with the reference simulations as well as with each other illustrating the the thermochemistry of the surface species is robust and can be used in future for similar investigations where thermodata is an important parameter.

The CO selectivity variation with temperature in methane steam reforming for fixed fuel ratio shown in Figure 6 shows deviations with thermodynamic model calculations. The calculations with LOGEcat using the kinetic model are in good agreement with the reference simulations and experimental data. Various reactions pathways and reaction sensitivity analysis needs to be performed in future to understand the difference in the calculations using the thermodynamic model.

To summarize, the 1D model, LOGEcat, is used to perform the simulations by utilizing already existing kinetic model with 42 irreversible reactions and with the help of new thermodynamic model consisting of 21 reversible reactions. In thermodynamic model, the reverse rates are calculated using the thermochemistry of the intermediate species. The thermochemistry of all the species is taken from [1] except for CO(s), which is taken from RMG [26] and adopted for our model by increasing the enthalpy of formation by 40 kJ to get a better agreement with the reference experiments and simulations.

4. Conclusions

The kinetic models are used to investigate the steam reforming of methane over a nickel catalyst. These models comprise of all the direct or irreversible reactions in the detailed surface reaction mechanism due to the unavailability of the thermochemistry of the intermediate species. For example, in Maier et al. [1], the reforming is studies by using 42 forward reactions. However, in the present work we focus on utilizing the thermochemistry of the surface bound species from different sources to develop a thermodynamic model.

Therefore, only 21 forward reactions are considered in the surface reaction mechanism and the backward reactions are omitted. The rate for backwards reactions are accounted by the thermochemistry of the intermediate species. In our model, the thermochemistry for the species CO(s) is taken from RMG [26] with a increased formation of enthalpy by 40 kJ and for all other species, the thermochemistry is utilized from [1]. A one-dimensional model, LOGEcat is used to perform the simulations. First, the kinetic model with 42 reactions from literature is used to carry out the simulations and then the thermodynamic model with 21 reactions. The results from both the models are compared with literature [1] for different temperatures.

The thermochemistry from different sources shows differences in the conversion of methane and water. The sensitivity analysis on the thermochemistry lead us to find the most sensitive species which can be modified to get the conversion of reactant comparable with the reference experiments and simulations performed with the kinetic model. A further investigation can be performed to find the sensitive limits for the formation of enthalpy of all the species considered to understand the role of thermochemistry in details. The most sensitive reactions can also be figured out by performing the flow analysis in future.

Acknowledgment
Financial support by the federal ministry of education and research (Bundesministerium für Bildung und Forschung, BMBF) under the Grant Number 03SF0693A of the collaborative research project "Energie-Innovationszentrum" is gratefully acknowledged.
The authors also thank the Graduate Research School (GRS) of the BTU Cottbus-Senftenberg for the partial travelling financial support.

References

[1] L. Maier, B. Schädel, K. H. Delgado, S. Tischer, and O. Deutschmann, "Steam Reforming of Methane Over Nickel: Development of a Multi-Step Surface Reaction Mechanism," *Topics in Catalysis.*, vol. 54, pp. 845–858, 2011.

[2] B. T. Schädel, M. Duisberg, and O. Deutschmann, "Steam reforming of methane, ethane, propane, butane, and natural gas over a rhodium-based catalyst," *Catalysis Today.*, vol. 142, pp. 42–51, 2009.

[3] J. R. Rostrup-Nielsen, *In: Anderson JR, Boudart M (eds) Catalytic steam reforming in catalysis—science and technology*. Berlin: Springer-Verlag, 1984.

[4] B. C. Michael, A. Donazzi, and L. D. Schmidt, "Effects of H2O and CO2 addition in catalytic partial oxidation of methane on Rh," *Journal of Catalysis.*, vol. 265, pp. 117–129, 2009.

[5] R. Quiceno, O. Deutschmann, J. Warnatz, and J. Pérez-Ramírez, "Modelling of the high-temperature catalytic partial oxidation of methane over platinum gauze. Detailed gas-phase and surface chemistries coupled with 3D flow field simulations," *Applied Catalysis A.*, vol. 303, pp. 166–176, 2006.

[6] J. Wei and E. Iglesia, "Isotopic and kinetic assessment of the mechanism of reactions of CH4 with CO2 or H2O to form synthesis gas and carbon on nickel catalysts," *Journal of Catalysis.*, vol. 224, pp. 370–383, 2004.

[7] L. M. Aparicio, "Transient Isotopic Studies and Microkinetic Modeling of Methane Reforming over Nickel Catalysts," *Journal of Catalysis.*, vol. 165, pp. 262–274, 1997.

[8] D. Dissanayake, M. P. Rosynek, K. C. C. Kharas, and J. H. Lunsford, "Partial Oxidation of Methane to Carbon Monoxide and Hydrogen over a Ni/Al2O3 Catalyst," *Catalysis Today.*, vol. 132, pp. 117–127, 1991.

[9] W. J. M. Vermeiren, E. Blomsma, and P. A. Jacobs, "Catalytic and thermodynamic approach of the oxyreforming reaction of methane," *Catalysis Today.*, vol. 13, pp. 427–436, 1992.

[10] S. Hannemann, J. D. Grunwaldt, N. van Vegten, A. Baiker, P. Boye, and C. G. Schroer, "Distinct Spatial Changes of the Catalyst Structure inside a Fixed-Bed Microreactor during the Partial Oxidation of Methane over Rh/Al2O3," *Catalysis Today.*, vol. 126, p. 54, 2007.

[11] D. A. Hickman and L. D. Schmidt, "Steps in CH4 oxidation on Pt and Rh surfaces: High-temperature reactor simulations," *American Institute of Chemical Engineers AIChE.*, vol. 39, pp. 1164–1177, 1993.

[12] A. M. Gadalla and M. E. Sommer, "Carbon dioxide reforming of methane on nickel catalysts," *Chemical Engineering Science.*, vol. 44, pp. 2825–2829, 1989.

[13] Z. W. Liu, H. S. Roh, and K. W. Jun, "Important factors on carbon dioxide reforming of methane over nickel-based catalysts," *Journal of Industrial and Engineering Chemistry.*, vol. 9, pp. 753–761, 2003.

[14] M. C. J. Bradford and M. A. Vannice, "Catalytic reforming of methane with carbon dioxide over nickel catalysts II. Reaction kinetics," *Applied Catalysis A: General.*, vol. 142, pp. 97–122, 1996.

[15] R. Schwiedernoch, S. Tischer, C. Correa, and O. Deutschmann, "Experimental and Numerical Study of the Transient Behavior of a Catalytic Partial Oxidation Monolith," *Chemical Engineering Science.*, vol. 58(3), pp. 633–642, 2003.

[16] O. Deutschmann and L. Schmidt, "Modeling the partial oxidation of methane in a short-contact-time reactor," *American Institute of Chemical Engineers AIChE.*, vol. 44, pp. 2465–2477, 1998.

[17] A. B. Mhadeshwar and D. G. J. Vlachos, "Hierarchical Multiscale Mechanism Development for Methane Partial Oxidation and Reforming and for Thermal Decomposition of Oxygenates on Rh," *The Journal of Physical Chemistry: B.*, vol. 109(35), pp. 16819–16835, 2005.

[18] J. Xu and G. F. Froment, "Methane steam reforming, methanation and water-gas shift: I. Intrinsic kinetics," *American Institute of Chemical Engineers AIChE.*, vol. 35, pp. 88–96, 1989.

[19] J. R. Rostrup-Nielsen and J. H. B. Hansen, "CO2-Reforming of Methane over Transition Metals," *Journal of Catalysis.*, vol. 144, pp. 38–49, 1993.

[20] "LOGEsoft v1.10." www.logesoft.com, 2008.

[21] Rakhi., V. Günther, J. Richter, and F. Mauss, "Steam reforming of methane over nickel catalyst using a one-dimensional model," *International Journal of Environmental Science.*, vol. 5(1), pp. 1–32, 2022.

[22] Rakhi, K. P. Shrestha, V. Günther, and F. Mauss, "Kinetically consistent detailed surface reaction mechanism for steam reforming of methane over nickel catalyst," *Reaction Kinetics, Mechanisms and Catalysis.*, 2022.

[23] Rakhi, V. Günther, and F. Mauss, "A detailed surface reaction mechanims to investiagate oxidation of methane over nickel catalyst," *Proceedings in Applied Mathematics and Mechanics.*, 2022.

[24] Rakhi, V. Günther, J. Richter, and F. Mauss, "Steam reforming of methane over a nickel-based catalyst," *In Linköping Electronic Conference Proceedings.*, 2022.

[25] Rakhi, V. Günther, and F. Mauss, "Insights into dry reforming of methane over nickel catalyst using a thermodynamic model," *Reaction Kinetics, Mechanisms and Catalysis.*, 2023.

[26] M. Liu, A. G. Dana, M. S. Johnson, M. J. Goldman, A. Jocher, A. M. Payne, C. A. Grambow, K. Han, N. W. Yee, E. J. Mazeau, K. Blondal, R. H. West, C. F. Goldsmith, and W. H. Green, "Reaction Mechanism Generator v3.0: Advances in Automatic Mechanism Generation," *Journal of Chemical Information and Modeling.*, vol. 61, pp. 2686–2696, 2021.

Developing a Multi-Building Scale Energy Model for a university campus using URBANopt

Hamed Mohseni Pahlavan[a,*], Natasa Nord[a]

[a] *Department of Energy and Process Technology, Norwegian University of Science and Technology (NTNU), Kolbjørn Hejes vei 1 B, Trondheim 7491, Norway*
hamed.m.pahlavan@ntnu.no

Abstract

Building energy models are developed to describe energy performance. The energy performance of buildings is influenced by physical and human influenced factors. Therefore, to improve energy efficiency and renewable energy implementation in buildings on large scale, there is a need to analyze buildings on a large scale. In this study, URBANopt, a multi-building energy evaluation tool, was used to develop an accurate Multi building scale energy model for a university campus. This model will be useful in the future work to evaluate various available and emerging building-level and district-level technologies and retrofitting options to improve energy performance. URBANopt is a unique tool that leverages high-fidelity simulations of buildings, community-scale systems, distributed energy resources, and the associated interactions with local distribution electric infrastructure. A university campus in Norway was chosen as a case study. Results obtained from URBANopt were compared with a typical building energy simulation model in IDA-ICE for a representative building. This representative building was developed based on building characteristics, functionality, and geographic location, including indoor and outdoor climate conditions. Both models were validated by using measurement data. The results showed better simulation accuracy of the multi-building method of URBANopt with the measurement data, mainly due to the averaging of the characteristics of all buildings in the development of the representative building. Furthermore, the URBANopt allowed assigning a different scenario of technologies and retrofit options to each building in the evaluation process, which is impossible in the typical model due to its nature. However, it should be pointed out that the computational time of the model developed in URBANopt was higher and will increase more with the increased number of buildings.

1. Introduction

The rapid urbanization has led to an increasing demand for energy in urban areas. The building sector consumes a significant portion of the energy used in urban areas, accounting for 30-40% of global energy use (Li, Zhou et al. 2017). As a result, modeling building energy use at an urban scale has become a crucial task in achieving energy efficiency in urban areas. Urban building energy modeling (UBEM) refers to the process of predicting urban building energy use using computer simulations. UBEM is an essential tool for predicting energy use and evaluating energy efficiency strategies in urban policy (Wang, Ferrando et al. 2022).

In order to create a reliable building energy model of a new or existing neighborhood, the task can be broken into the following subtasks: simulation input organization (data input), thermal model generation and execution (thermal modeling) as well as result validation (validation) (Reinhart and Cerezo Davila 2016). The simulation input organization is concerned with the collection and integration of data from various sources, such as weather data, building design data, and energy use data, to create a comprehensive and accurate input dataset for the model (Wang, Ferrando et al. 2022). Once the input data has been collected and integrated, the thermal model generation and execution stage involves creating a mathematical model of the energy use of the buildings in the urban area, which can then be simulated and tested under various conditions (Wang, Ferrando et al. 2022).

Several different types of energy models have been proposed for modeling urban building energy use over the past few decades, each with their own strengths and weaknesses (Li, Zhou et al. 2017). Physics-based, bottom-up models are one of the most common types of models used for this purpose. These models rely on detailed physical data, such as the building's geometry, construction materials, and HVAC system, to generate a comprehensive model of the energy use of the building. These models are typically accurate but can be time-consuming to develop and require a lot of detailed data (Li, Zhou et al. 2017).

Another approach to modeling urban building energy use involves coupling bottom-up physics models with geographic information systems (GIS) techniques. This approach involves using GIS to integrate the physical data of the building with the spatial data of the urban area to create a more comprehensive model. One study that used this approach modeled urban building energy use and CO_2 emissions for Indianapolis-Marion County, IN by integrating their energy use model, eQUEST, with GIS techniques (Li, Zhou et al. 2017).

In addition to physics-based models, statistical models have also been used for modeling urban building energy use. These models rely on statistical analysis of data to generate a model of the energy use of the buildings in the urban area. While these models are typically faster and require less detailed data than physics-based models, they are also generally less accurate (Li, Zhou et al. 2017).

One important aspect of modeling building energy use at an urban scale is the development of archetype libraries. Archetype libraries are collections of building models that have been grouped into homogenous groups based on their characteristics, such as building type, size, and construction materials. These libraries can be used to streamline the modeling process by providing pre-existing models that can be easily adapted to new urban areas (Mohammadiziazi, Copeland et al. 2021).

The benefits of modeling building energy use at an urban scale are numerous. By accurately modeling the energy use of buildings in an urban area, policymakers and energy planners can identify areas of high energy use and develop targeted strategies for improving energy efficiency. This can lead to reduced energy use, lower energy costs, and reduced greenhouse gas emissions. In addition, modeling building energy use at an urban scale can also help to identify areas of the urban environment that are particularly vulnerable to heat waves and other extreme weather events and validate the performance of UBEMs. These validation techniques can range from comparing the simulated energy use with measured energy use data, to comparing the simulated thermal loads with real weather data (Li, Zhou et al. 2017).

One interesting application of UBEMs is the study of waste heat from buildings and its contribution to urban heat islands. A study conducted by the US Department of Energy found that during heat waves, waste heat from air conditioning can increase the amount of heat being dispersed from buildings to the urban environment by up to 20% (Luo, Vahmani et al. 2020). UBEMs can be used to simulate the impact of waste heat on urban temperatures, which can help policymakers develop strategies to reduce urban heat island effects.

Detailed building energy data and existing buildings that match prototypical building energy models (BEMs) are essential factors for the current UBEM development. There are several instances where building energy data are unavailable due to privacy concerns, lack of civic energy disclosure requirements, or properties that do not meet reporting threshold requirements. As highlighted by different studies (Holloway and Bunker 2006, Abrahamse and Steg 2009, Chen, Xu et al. 2017, Chen, Feng et al. 2022), establishing a correlation between energy usage and factors like socioeconomic status, climate, and building characteristics has been challenging. This may result in significant differences between a model based solely on prototype BEMs and a real community. This paper investigated the situation where energy data is only available at the aggregate district level, and current prototype BEMs do not account for actual energy usage in the community. This work aimed to establish a pathway for precise district-level building energy model creation with limited data.

Based on literature, UBEM, while valuable, has several limitations, three important ones are:

Data Availability: Gathering comprehensive and up-to-date data for large-scale urban models can be challenging. Limited data can lead to less accurate simulations.

Complexity: Urban environments are complex and dynamic, making it difficult to capture all factors affecting energy use accurately. This complexity can lead to simplified models that may not represent reality well.

Computational Intensity: Simulating energy use in large urban areas requires substantial computational resources, and it can be time-consuming. This limits the ability to perform real-time simulations or analyze numerous scenarios quickly.

The existing literature may be enriched by the present work, which illustrates how to create a reliable multi-building-scale energy model (MBSEM) for an untypical district when detailed energy data are unavailable. By highlighting essential datasets, tools, and partnerships, this addition establishes a roadmap for developing a model on a district scale. This contribution is accomplished through a university campus case study in Trondheim, Norway. The Urban scale energy simulation tool URBANopt (Polly, Chuck Kutscher et al. 2016, Kontar, Ben Polly et al. 2020) was used. The current effort is focused on describing the model development process.

2. Methodology

This section describes the case study, the URBANopt tool, and the MBSEM development.

2.1. Gløshaugen campus

The case study case in this paper is a university campus located in Trondheim, Norway. In the Gløshaugen campus, the system supplies heat to a total building area of 300,000 m², and the main functions of these buildings are education, offices, laboratories, and sports. The campus district heating (DH) system is connected to the city DH system by the main substation (MS). Apart from the heat supply from the city DH system, part of the annual heat supply comes from waste heat recovered from the university's data center (DC) (Li, Hou et al. 2021). According to the measurements from June 2017 to May 2018, the total heat supply for the campus DH system was 32.8 GWh. About 80% of the heat supply came from the central DH system through the MS. The other 20% came from the waste heat recovery from the DC (Li 2022). Map of the Gløshaugen campus is shown in Figure 2.

2.2. URBANopt tool

URBANopt is a physics-based energy modeling platform for districts and communities (Polly, Chuck Kutscher et al. 2016, Kontar, Ben Polly et al. 2020). URBANopt is a modular, open-source SDK, built on DOE tools such as EnergyPlus, OpenStudio, and Spawn of EnergyPlus. URBANopt includes capabilities and workflows that enable multi-building analysis at a neighborhood, district, or campus scale (generally 10 s to 100 s of buildings) and connections to other tools and engines that allow for the analysis of shared energy systems, distributed energy resources (DER), and the electric distribution systems, including interactions and impacts with building efficiency and demand flexibility strategies (Laboratory, 2022). Figure 1 (Fallahi, Sammy Houssainy et al. 2022) shows the structure of different tools in URBANOpt SDK.

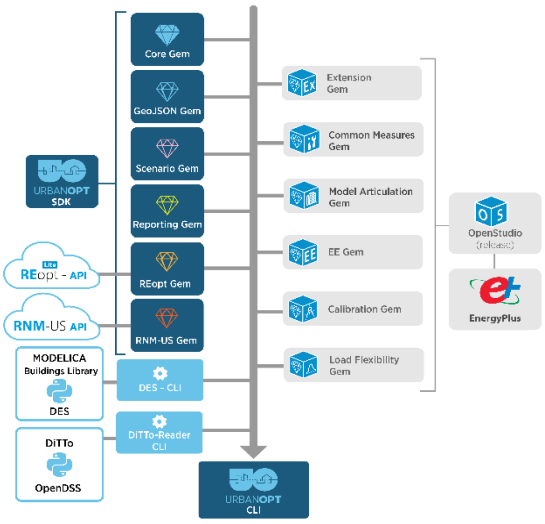


Figure 1. URBANopt SDK Gem structure

URBANopt helps manage geospatial information for modeling a community and automates the creation of detailed physics-based models for baseline scenarios (e.g., existing conditions) and advanced performance scenarios (e.g., retrofit upgrades). It exchanges data with other tools, manages simulations, and evaluates and compares scenarios. In this study one GeoJSON file describes Gløshaugen campus buildings, energy systems, and end uses, one CSV file tunes and implements scenarios, and another CSV file links building models to scenarios. URBANopt workflows for generating commercial building models were described in (Kontar, Ben Polly et al. 2020, Charan, Mackey et al. 2021).

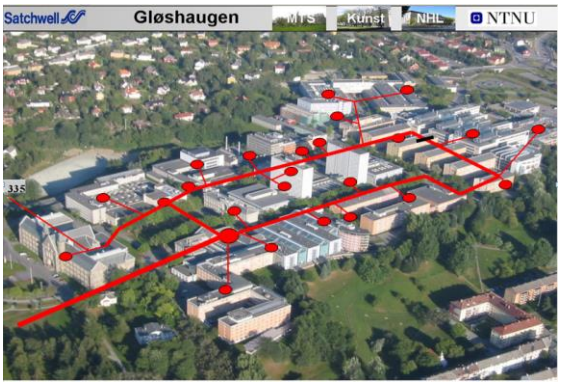


Figure 2. Aerial image and building cluster for the Gløshaugen campus.

2.3. Multi-building scale energy model development

This section describes the process of developing the Gløshaugen campus MBSEM, emphasizing critical datasets and resources. The objective of model development is to generate a precise and physics-based representation of energy use in the district. The first critical dataset is actual utility usage data for the district. There is a dataset for the campus DH demand for 2017, as shown in Figure 3.

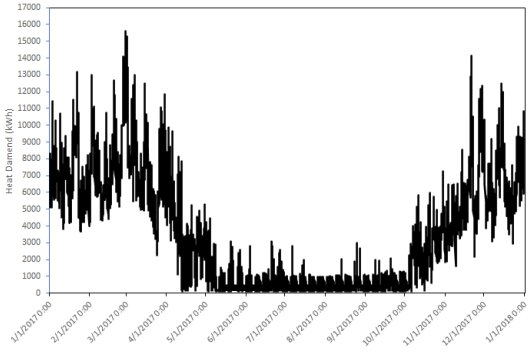


Figure 3. Hourly buildings heat demand for the year 2017

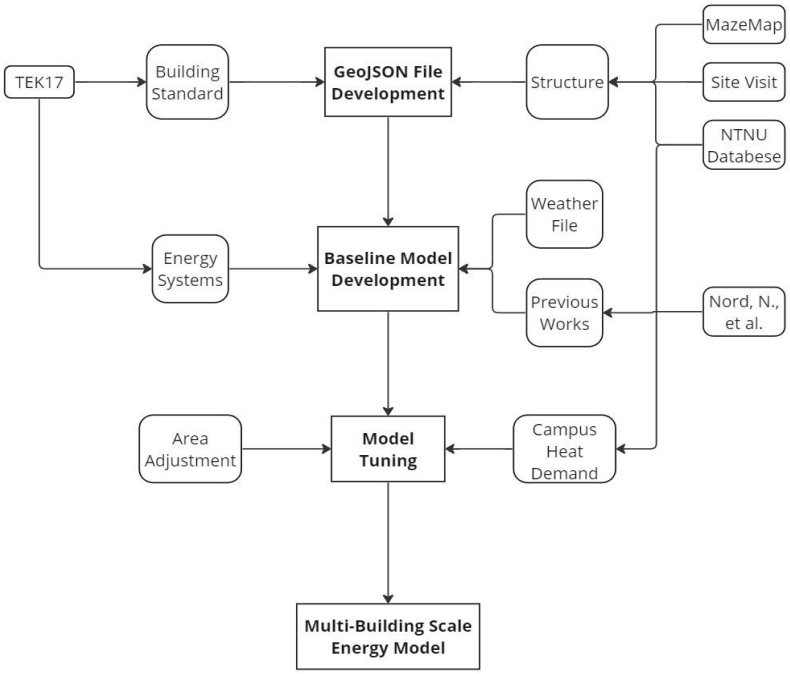


Figure 4. Summary of model development and tuning process, including data sources used to develop simulation result targets.

The assembly of the MBSEM was carried out in six steps. First, GeoJSON format was used to define building geometry, construction sets, and energy systems, which were then used for the URBANopt modeling platform. Second, OpenStudio measures were utilized to define building energy systems. Examples of these included HVAC, hydronic heating system types, and component efficiencies. Third step was the development of baseline predictions for building heat demand. In the fourth step, heat demands were adjusted for BEM end-users developed in step 3. The fifth step looked at BEM heat demands and compared to DH data shown in Figure 3. Finally, errors were found and the MBSEM was modified in the sixth step. This workflow is presented graphically in Figure 4, showing data sources and inputs. The different data used to develop the tuned MBSEM presented in this work can be broken down into three categories:

1. Building structure: Geometry (location, area, no. of stories etc.) and materials used in walls, windows, and building exterior.
2. Building energy systems: The performance characteristics of end use energy systems for heating, such as HVAC, domestic hot water (DHW), and all other devices/systems/appliances powered by district heating.
3. Aggregated building heat demand: Patterns that define how and when heat is needed on campus.

2.3.1. Building structure

Building geometries were developed using aerial images, Mazemap[1], and site visits. This information along with some building regulations, were used as input to develop the GeoJSON file as the first step of model development (Figure 4). The GeoJSON format was used to describe geometries. Building stories were determined using Mazemap and site visits. The campus buildings arrangement is displayed in Figure 5, modeled in GeoJSON format. The two assumptions were made for geometry development in buildings with mixed-use and asymmetric floor areas in some multi-story buildings. First, the current building workflow did not account for mixed-use buildings. In these instances, new uses were added to the library of the tool. For buildings with asymmetric floor areas by story, these buildings were split and modeled as separate buildings with symmetric floor areas.

Figure 5. Gløshaugen campus (GeoJSON format)

[1] MazeMap Indoor Navigation App. MazeMap Indoor Navigation. (n.d.). https://link.mazemap.com/fSPmwDL1

2.3.2. Building energy systems

Building energy systems address the properties of space conditioning and domestic hot water (DHW) systems. Trondheim, where the case study is located, has a Nordic climate, meaning space heating accounts for the majority of energy usage in these buildings. Therefore, the focus of this study was on heating demand and energy systems associated with it. Energy systems data, weather data, and other building information were used as inputs to create the baseline model of the campus from the pre-developed GeoJSON file, as shown in Figure 4.

2.3.3. Aggregated building heat demand

The entire campus's aggregated heat demand was utilized as input for model tuning and validation. Hourly space heating demand and DHW usage over one year make up this data.

2.3.4. Baseline load

Collected data on building structures and energy systems described in Sections 2.3.1 and 2.3.2 were converted to a GeoJSON format used by URBANopt to create a base model. Some key parameters defined in (Nord, Sandberg et al. 2019) were used for creating the base model. These input parameters included characterization of the building envelope, occupancy behavior, and building functionality. This initial model was used to create a baseline heat demand of the campus as an input in the tuning process of the model (3rd step in Figure 4).

2.3.5. Baseline tuning

OpenStudio measures were applied to tune URBANopt simulation results to match the baseline load described in Section 2.3.4 with aggregate heat demand described in Section 2.3.3. TMY3 weather data files for Trondheim, Norway, were used in this work.

Tuning occurred in two steps (model tuning part in Figure 4). First, total conditioned floor area was tuned to match heat demand values. After area tuning, the complete district-scale energy model was benchmarked against the measured data shown in Figure 3 by adjusting the model thermostat values.

For example, in one part of the tuning process, the initial thermostat set point for the model was a variable set point. 21°C for working hours and 15°C for non-working hours. The initial results showed a large difference between the simulation output and the measured data, especially in non-working hours. For this reason, each time, by increasing the temperature of the thermostat during non-working hours in the model with a step of one degree, the output of the model and the measured data were compared. Finally, the set point temperature of 19°C for non-working hours achieved the best adaptation in the results.

3. MBSEM simulation results

Results from the MBSEM are presented in two sections. In the first section, the accuracy of the simulation is measured by comparing the results with the actual heat demand data shown in Figure 3. The second section compares the heat demand produced in the MBSEM simulation against a typical building energy simulation model in the IDA-ICE tool, developed as a representative building for Gløshaugen campus (Nord, Sandberg et al. 2019). This representative building was developed based on building characteristics, functionality, and geographic location, including indoor and outdoor climate conditions.

3.1. Simulation accuracy

The evaluation of building energy models' accuracy is a necessary task, as it allows for the implementation and investigation of energy-saving strategies while maintaining human comfort. ASHRAE Guideline 14-2014, the International Performance Measurement and Verification Protocol (IPMVP), and the Federal Energy Management Program (FEMP) are the most widely recognized methodologies for evaluating the accuracy of these models (Ruiz and Bandera 2017). Normalized Mean Bias Error (NMBE), Coefficient of Variation of the Root Mean Square Error (CV(RMSE)), and coefficient of determination (R^2) are the principal accuracy indices used in these standards. This study used NMBE and CV(RMSE) as error indicators for our simulation results.

Table 1 shows the total annual heat demand based on measured data and simulation outputs. There is a 4 GWh underestimation in simulation results for the total heat demand on campus. Table 2 shows the accuracy indicators on an hourly and monthly basis. According to ASHRAE Guideline 14, the acceptance criteria for these indicators are +5% NMBE and 15% CV(RMSE) for hourly data and +10% NMBE and 30% CV(RMSE) for monthly data. Therefore, the simulation results could be acceptable, considering that the output is heat demand for a MBSEM instead of an individual BEM.

Table 1. Annual heat demand

	Heat Demand
Actual	32.8 GWh
Simulation	28.8 GWh

Table 2. Simulation error

Data type	**MBE**	**CV (RMSE)**
Hourly	8%	35%
Monthly	8%	12%

Figure 6 shows actual and simulation heat demand for a typical winter week starting from Saturday. The simulation provided an accurate forecast of hourly heat demand for both weekdays and weekends. The model's heat demand was at a minimum value during the night on working days, resulting in a significant difference. This also led to high peaks at the start of the working hours. Tuning process can reduce the difference by adjusting the thermostat of each building.

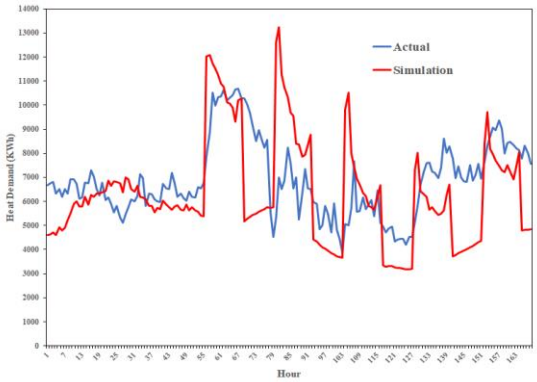


Figure 6. Heat demand for a typical winter week (starting from Saturday)

Figure 7 shows the monthly heat demand for both measured and simulation data. According to this, the underestimation in heat demand can also be seen in monthly data. The underestimation was more common in the cold season, when the demand for heat came mainly from space heating. However, during summer, when heat demand is solely for DHW, the developed model showed better prediction.as a result, an alternative approach may be necessary for adjusting thermostats despite sufficient input data for building energy systems.

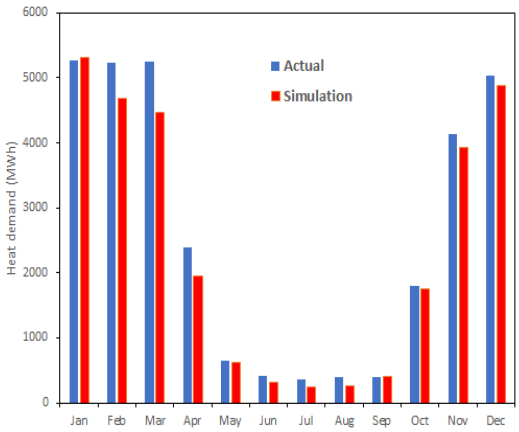


Figure 7. A comparison of actual monthly heat demand versus simulated data

3.2. MBSEM and IDA ICE typical building comparison

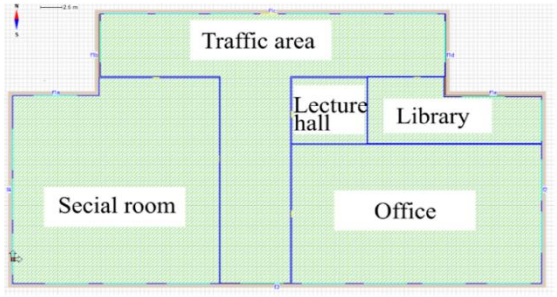
Using typical/reference buildings is another approach to deal with district-level building energy analysis. The most frequent building design in the examined area forms the foundation of the district's reference model. For the Gløshaugen campus, based on the average geometry, building envelope parameters, occupancy behavior, etc., a reference model was built in IDA-ICE simulation software (Nord, Sandberg et al. 2019).

To create the IDA-ICe model, information about the distribution of areas and rooms was provided by the Technical Management Section at NTNU.. It was found that the total area was divided by 140 rooms and 18 zones. Eventually, all zones have been combined to form the nine most representative: office, reading hall, lecture hall, laboratory, traffic area, technical room, workshop, cleaning and sanitary room and other. some zones with similar functionality were combined for creating the model and finally, the geometry and size have been selected for reference building. Table 3 summarizes key information of the Reference model building areas. The simulation model and the floor area distribution are shown in Figure 8.

Table 3. Reference model building areas

Building geometry	**Parameter**	**Reference model**
General	Total area [m²]	7220.00
	Heated are gross [m²]	7159.20
	Floor area [m²]	1805.00
	Number of floors	4
Total zone area/ per floor area	Office [m²]	1967.60 / 491.90
	Library [m²]	545.20 / 136.30
	Educational facilities [m²]	282.00 / 70.50
	Special room [m²]	2321.20 / 580.30
	Traffic area [m²]	2043.20 / 510.80

(a) Simulation model



(b) Floor area

Figure 8. Simulation model developed in IDA-ICE

Based on the geometry and building envelope parameters, the model was built in IDA-ICE simulation software. Building envelope parameters and other important values were defined as weighted averages and shown in Table 4.

Table 4. building envelope parameters for IDA-ICE model

Category	Parameter	Reference model
U-value	External wall [W/m^{2}K]	0.57
	Internal wall [W/m^{2}K]	0.62
	External floor [W/m^{2}K]	0.19
	Internal floor [W/m^{2}K]	2.39
	Windows [W/m^{2}K]	2.19
	Doors [W/m^{2}K]	1.09
	Roof [W/m^{2}K]	0.48
General for façade	Normalized thermal bridge value [W/m^{2}K]	0.10
	Infiltration [l/h]	3.07
	Total windows area [%]	13.16

Since the conditioned floor area of this building is about 7 200 m², an area adjustment took place to calculate the total heat demand of the campus, based on this model. After this adjustment, the total annual heat demand of the Gløshaugen campus was 25.5 GWh (7.2 GWh less than the measured data). The monthly heat demand of the campus, based on IDA ICE, MBSEM, and measured data, is shown in Figure 9.

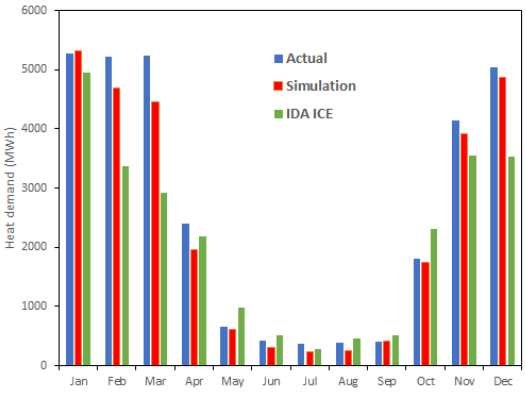


Figure 9. A comparison of actual monthly heat demand versus simulated data and IDA ICE results

Despite the MBSEM, the IDA ICE model overestimated the heat demand during the summer. During the cold season, especially in December, February, and March, there was a significant difference between IDA ICE results and measured data. The reason for this could be less heat loss through building envelope by aggregating all buildings in one typical building. This is something that should be considered in developing representative buildings.

This work aimed to demonstrate a replicable and scalable method for simulating multi-building scale energy models with minimum data available. Results from the URBANopt simulation tool showed good accuracy with measured heat demand. In general, the developed model underestimated the heat demand for campus (could be obtained from Table 1 and Figure 7). Adding more details of buildings' structure and energy systems and better adjustment in the tunning process could result in better accuracy of developed MBSEM. Furthermore, results obtained from the typical model showed even more underestimation of the heat demand of the Gløshaugen campus. This was mainly due to neglecting some properties of buildings in averaging process. However, it should be pointed out that the computational time of the model developed in URBANopt was higher and will increase more with the increased number of buildings, especially in the first stage, which is generating building energy models.

5. Summary and conclusions

This paper presents a process for developing a MBSEM for a university campus in Trondheim, Norway. The modeled district includes 24 educational, office, and laboratory buildings. The district-scale energy model includes individual BEM for all buildings. MBSEM tuning is accomplished through the matching heat demand values using local datasets. The developed model had acceptable accuracy on both monthly and hourly basis. MBSEM compared with a typical BEM developed for the same case study. The results showed better simulation accuracy of the MBSEM compared to typical BEM, mainly due to the averaging of the characteristics of all buildings in the development of this model. The approach facilitates detailed load construction to prepare for analyzing energy efficiency measures, electrification, onsite renewable energy conversion, and storage technologies.

References

Abrahamse, W. and L. Steg (2009). "How do socio-demographic and psychological factors relate to households' direct and indirect energy use and savings?" Journal of Economic Psychology **30**(5): 711-720.

Charan, T., et al. (2021). "Integration of Open-Source URBANopt and Dragonfly Energy Modeling Capabilities into Practitioner Workflows for District-Scale Planning and Design." Energies **14**(18).

Chen, C.-f., et al. (2017). "Thermal comfort or money saving? Exploring intentions to conserve energy among low-income households in the United States." Energy Research & Social Science **26**: 61-71.

Chen, C. F., et al. (2022). "Localized energy burden, concentrated disadvantage, and the feminization of energy poverty." iScience **25**(4): 104139.

Fallahi, Z., et al. (2022). Designing for Zero Energy and Zero Carbon on a Multi-Building Scale using URBANopt. Building Performance Analysis Conference and SimBuild (BPAC).

Holloway, D. and R. Bunker (2006). "Planning, Housing and Energy Use: A Review." Urban Policy and Research **24**(1): 115-126.

Kontar, R. E., et al. (2020). URBANopt: An Open-Source Software Development Kit for Community and Urban District Energy Modeling. Building Performance Analysis Conference and SimBuild (BPAC).

Li, H. (2022). Economic optimization for heatprosumer-based district heating systems in unidirectional heating markets.

Li, H., et al. (2021). "Energy, economic, and environmental analysis of integration of thermal energy storage into district heating systems using waste heat from data centres." Energy **219**.

Li, W., et al. (2017). "Modeling urban building energy use: A review of modeling approaches and procedures." Energy **141**: 2445-2457.

Luo, X., et al. (2020). "City-Scale Building Anthropogenic Heating during Heat Waves." Atmosphere **11**(11).

Mohammadiziazi, R., et al. (2021). "Urban building energy model: Database development, validation, and application for commercial building stock." Energy and Buildings **248**.

Nord, N., et al. (2019). Future energy pathways for a university campus considering possibilities for energy efficiency improvements. IOP Conference Series: Earth and Environmental Science, IOP Publishing.

Polly, B., et al. (2016). From Zero Energy Buildings to Zero Energy Districts Summer Study on Energy Efficiency in Buildings.

Reinhart, C. F. and C. Cerezo Davila (2016). "Urban building energy modeling – A review of a nascent field." Building and Environment **97**: 196-202.

Ruiz, G. and C. Bandera (2017). "Validation of Calibrated Energy Models: Common Errors." Energies **10**(10).

Wang, C., et al. (2022). "Data acquisition for urban building energy modeling: A review." Building and Environment **217**.

Modeling and control of WRRF biogas production

Tiina M. Komulainen[a,*] Bilal Mukhtar[a] Truls Ødegaard[a] Hilde Johansen[b] Kristine Haualand[b] Kjell Rune Jonassen[b] Simen Antonsen[a]

[a]*Oslo Metropolitan University, Norway,* [b]*Veas, Norway*
[*]tiina.komulainen@oslomet.no

Abstract

Wastewater treatment sector uses about 1 percent of total energy consumption in European Union, hence development of energy-efficient digital technologies is an urgent challenge. The aim of this article is to develop energy-efficient control strategies for biogas production from sewage sludge at water resource recovery facilities (WRRF). The case study is developed in collaboration Veas WRRF, Norway. The Veas biogas plant is operated semi-continuously in mesophilic conditions. The process includes inlet sludge pumps, four anaerobic digesters, heat exchangers for sludge heating, pumps for sludge recirculation and a compressor for gas recirculation. The process has two controlled variables, biogas flowrate and digester temperature, the main disturbance is the inlet substrate composition. The manipulated variables are flowrates of the inlet sludge, heating medium, and sludge recirculation. The real semi-continuous operation approximated as continuous operation with two hour moving averaging. Transfer functions were identified from the pre-processed data. The accuracy of the models was sufficient $14-60\%$. The transfer functions were used to design control strategies with PID-controllers and model predictive controller (MPC). The results show that both control strategies can increase biogas production and decrease variability in controlled and manipulated variables compared to the plant operation. MPC gave the best results, increasing biogas production up to 10 % and decreasing variability in controlled variables by $50-80\%$ and by $92-99\%$ in manipulated variables. These results indicate that implementation of advanced control technologies can improve the energy efficiency of biogas production.

1 Introduction

Wastewater treatment sector uses about 1 percent of total energy consumption in European Union, generating a high energy bill covered by the taxpayers (EuropeanCommission, 2022). The EU has set a goal for energy-neutrality in the wastewater sector by 2040 with renewable energy production, carbon neutrality and a resource-efficient bioeconomy (EuropeanCommission, 2021b) (EuropeanCommission, 2021a). Norwegian wastewater industries have ambitious targets to reduce the environmental impact (NorskVann, 2017), therefore, investment in energy-efficient biogas production is essential. Biogas can replace diesel and other fossil energy carriers in transport industry increasing the income for WRRFs and reducing the environmental impacts associated with biogas production and use. In our previous work we have reviewed digital technologies that can improve energy-efficiency in water industry to meet these demands (Komulainen & Johansen, 2021).

Several complex models have been suggested and applied for biogas production from sewage sludge. Veas WRRF biogas production has previously been modeled with the anaerobic digestion model nr 1, ADM1, with complex influent characterization by (Bergland & Bakke, 2016). (Attar & Haugen, 2019) have continued the work adapting to a simpler AM2 model with two substrate types. The simplest first principles modeling approach is continuous stirred tank reactor, a chemostat, with only one substrate type (Seborg et al., 2017). In recent master thesis work in collaboration with Veas, Mukhtar (2023) used the AM2 model to identify transfer function models for the biogas production, and plant data to identify models of the heat exchanger and pumps. Using these transfer function models, Mukhtar (2023) developed PID and MPC control algorithms for the Veas biogas process.

As VEAS biogas process has only one online process measurements related to the substrate, total suspended solids, in this article, we continue Mukhtar's work by identifying transfer function models of the anaerobic digestion process directly from plant data. Further, we use the new transfer function models to develop and test energy-efficient control strategies. Our research question is "Which control algorithms can optimize energy consumption and maximize biogas production?"

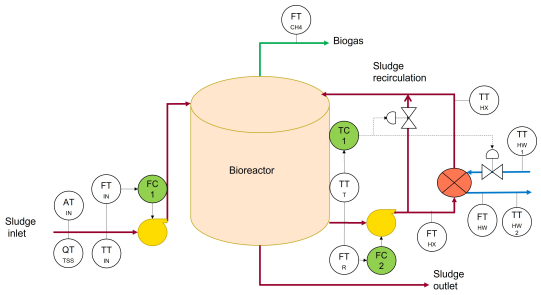


Figure 1. Simplified Veas biogas process with instrumentation

2 Materials and Methods

2.1 Software

Matlab software package version R2023a was used for the control experiments and System Identification Toolbox for estimation of the model parameters. Simulink solver algorithm was ode23s with automatic settings for the time step and error tolerance.

2.2 Biogas plant and instrumentation

A simplified illustration of the biogas plant and instrumentation is given in Figure 1. The Veas biogas plant has sequential operation with four bioreactors. The sequence for one bioreactor includes filling 5/6 of the tank with fresh sludge, heating the sludge until bioreactor temperature T reaches 37^oC. The anaerobic digestion process in mesophilic conditions is operated for about three weeks. The bioreactor is constantly fed with fresh sludge, flowrate F_{in} in and out of the bioreactor are the same, i.e. the bioreactor tank has constant hold-up. The sludge and biogas are semi-continuously recirculated in the bioreactor to avoid sedimentation. Measurements of all unit operations, except biogas recirculation rate are available, listed in Table 1.

2.3 Data collection and pre-processing

The online data set and laboratory data sets were collected for a period of one month 30.6.2022-30.7.2022. ABB Edge Insight was used to collect the online data from the SCADA system in .csv format. The laboratory data set was in .xlsx format.
The outliers in the online data set were first removed. Then, the missing values in the online data sets were filled.

2.4 Modeling

Transfer functions can be used for simplified modeling and control strategy design. These linear models can be developed following system identification procedure by (Ljung, 1999). The relationship between

Table 1. Online measurements

Symbol	Description	Unit
F_{in}	Flowrate sludge inlet	m^3/h
TSS_{in}	Total suspended solids in sludge at inlet	g/m^3
RPM_{in}	Pump speed inlet	rpm
T_{in}	Temperature sludge inlet	^oC
T_{HX}	Temperature sludge after HX	^oC
T	Temperature bioreactor	^oC
F_{HW}	Flowrate hot water in	m^3/h
T_{HW1}	Temperature hot water in	^oC
T_{HW2}	Temperature hot water out	^oC
R	Flowrate sludge recirculation	m^3/h
F_{HX}	Flowrate sludge via heat exchanger	m^3/h
RPM_R	Recycle pump speed	rpm
F_{CH4}	Biogas out	m^3/h

input variables $U_i(t)$ and output variable $Y(t)$ are assumed to be first order models with gain (K_p), time constant (T_p) and delay (T_d), presented in Equation 1:

$$TF(s) = \frac{Y(s)}{U_i(s)} = \frac{K_p}{(T_p s + 1)} e^{-T_d s} \quad (1)$$

2.5 Control methods

The transfer function models identified from Veas data were used for parametrization and tuning of the PID controllers and the MPC controllers. Tuning rules were adapted from (Skogestad, 2003) and (Seborg et al., 2017).

2.6 Error indices

The data-driven models are compared with each other using the fitness index (FIT) and integral of absolute error (IAE) between the real measurements $y_{i,measurement}$ and the model calculated output $y_{i,model}$ over N samples. The fitness index is calculated with Equation 2, where norm is the Euclidean norm.

$$FIT = (1 - \frac{norm(y_{i,meas} - y_{i,model})}{norm(y_{i,meas} - y_{i,mean})})100 \quad (2)$$

$$IAE = \int_0^N |y_{i,meas}(t) - y_{i,model}(t)|\,dt \quad (3)$$

The control results are evaluated using the integral of absolute error between the setpoint and measured value

$$IAE = \int_0^N |y_{sp}(t) - y(t)|\,dt \quad (4)$$

and integral of absolute movement in manipulated variables:

$$IAMV = \int_0^N |u_i(t) - u_i(t-1)|\,dt \quad (5)$$

Table 2. Mean, standard deviation and applied scaling of process variables.

Variable	Mean	St.Dev	Scaling
F_{IN}	7.049	0.8763	120 min MA
F_{HX}	42.2324	0.0541	120 min MA
T_{IN}	21.4041	0.8398	120 min MA
T_{HX}	38.3261	0.5049	120 min MA
TSS_{IN}	6.7175	0.7719	120 min MA
F_{HW}	12.2306	7.7764	120 min MA
T_{HW}	52.9991	0.8505	120 min MA
F_{CH4}	123.9211	13.7746	120 min MA
T	37.0662	0.0424	120 min MA

3 Results

3.1 Data description

For the modeling, online data from Veas WRRF were used for period 30.6.2022-30.7.2022 with 10 minutes sampling time. First half of the data was used for estimation and second half for validation of the transfer function models. As the size of the data set is limited, the same data set was used for the control experiments. The process variables with mean and standard deviation are presented in Table 2.

3.2 Data pre-processing

The outliers in the data set were identified based on the three standard deviation rule, and removed. Then, the missing values in the data sets were filled in using Matlab knnimpute function based on nearest-neighbor imputation method. The inlet temperature sensor T_{IN} is placed into a joint pipeline between inlet sludge and recirculated sludge, where recirculation is on 30 minutes and off 30 minutes. Hence, for T_{IN} before the moving averaging, the inlet temperature values over 20 ^oC were removed and replaced with previous temperature value under 20 ^oC.

Due to the sequential operation of the Veas biogas plant, all the variables exhibit high variation. The sampling interval is 10 minutes. Different moving average window sizes were tested, but considering the process time constants of 300-2800 minutes, 120 minutes (12 samples) window was chosen as a window of 60 minutes did not decrease significantly the high variation in the raw data.
A moving average of 120 minutes was applied to all input and output data. Without the moving average, the system identification did not work properly. Then, the mean values, given in Table 2, were removed.

3.3 Modeling

For control purposes, the process was to be modeled using transfer functions identified from plant data. To allow control strategy design, the process was divided into three subprocesses with one output variable each. The controlled variables are (1) biogas flowrate F_{CH4} out of the bioreactor, (2) temperature T in bioreactor and (3) recycled sludge outlet temperature T_{HX} after a heat exchanger. Different combinations of input variables were tested for the subprocesses. Some of the input variables were omitted if the parameter uncertainty got very high (thousand times larger than the parameter value) or if the time constant was very high (many thousands of minutes). For example, recycle rate R, is dependent on heated sludge flow rate F_{HX}, and omitting R improved the modeling results.

The biogas production F_{CH4} in the bioreactor is dependent on inlet sludge flowrate F, inlet suspended solids percentage TSS_{in}, inlet temperature T_{in} and recirculated and heated sludge flow rate F_{HX} and temperature T_{HX}. Surprisingly, inlet temperature T_{in} gave negative relationship to biogas production and was omitted as input variable. The model parameters and model fitness are given in Table 3. The model prediction and the measured value are illustrated in Figure 2. The model prediction is following the main trends of the biogas production. As an unmeasured part of the biogas is recirculated back to the digesters, the measured value has a rapid variation that the model cannot capture.

$$F_{CH4}(s) = TF_{11}(s)F(s) + TF_{12}(s)TSS_{in}(s) + TF_{13}(s)T_{in}(s) + TF_{14}(s)F_{HX}(s) + TF_{15}(s)T_{HX}(s) \quad (6)$$

Temperature in the bioreactor T is dependent on the same input variables, except inlet suspended solids percentage TSS_{in}. The model parameters and model fitness are given in Table 4. The model prediction and the measured value are illustrated in Figure 3. The model prediction follows the main trends in the measurement, but the rapid variation in the data was not captured.

$$T(s) = TF_{21}(s)F(s) + TF_{22}(s)TSS_{in}(s) + TF_{23}(s)T_{in}(s) + TF_{24}(s)F_{HX}(s) + TF_{25}(s)T_{HX}(s) \quad (7)$$

Further, the heat exchanger was modeled as first order transfer function between the HX outlet temperature T_{HX} and hot water variables F_{HW} and T_{HW1} and sludge flowrate F_{HX} and temperature $T(s)$. The model parameters and model fitness are given in Table 5.

The model prediction and the measured value are illustrated in Figure 4 . The model prediction follows the main trends in the measurement.

$$T_{HX}(s) = TF_{31}(s)F_{HW}(s) + TF_{32}(s)F_{HX}(s) + TF_{33}(s)T_{HW1}(s) + TF_{34}(s)T_{(}s) \quad (8)$$

Table 3. Transfer function parameters for between F_{CH4} and inputs.

Input	Kp [-]	Tp1 [min]	Td [min]
F_{in}	12.76	1350	0
TSS_{in}	11.80	101	167
Tin_{in}	0	0	0
F_{HX}	-21.15	29	266
T_{HX}	4.77	432	0
Error index	FITest %	FITval %	
	59.57	18.52	

Table 4. Transfer function parameters for bioreactor temperature T and inputs.

Input	Kp [-]	Tp1 [min]	Td [min]
F_{in}	$1.549 \cdot 10^{-}3$	0	298
TSS_{in}	0	0	0
Tin_{in}	$31.05 \cdot 10^{-}3$	853	233
F_{HX}	$-171.6 \cdot 10^{-}3$	20	87
T_{HX}	$111.9 \cdot 10^{-}3$	379	300
Error index	FITest %	FITval %	
	17.00	14.24	

Table 5. Transfer function parameters between heat exchanger outlet sludge temperature T_{HX} and inputs.

Input	Kp [-]	Tp1 [min]	Td [min]
F_{HX}	-0.19072	0	0
T	0	0	0
F_{HW}	0.059157	0	0
T_{HW1}	0.650889	0	0
Error index	FITest %	FITval %	
	47.04	48.27	

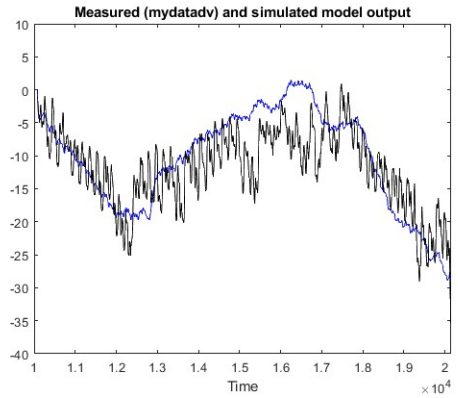


Figure 2. Scaled biogas flowrate F_{CH4} data (black) and model (blue), time in minutes.

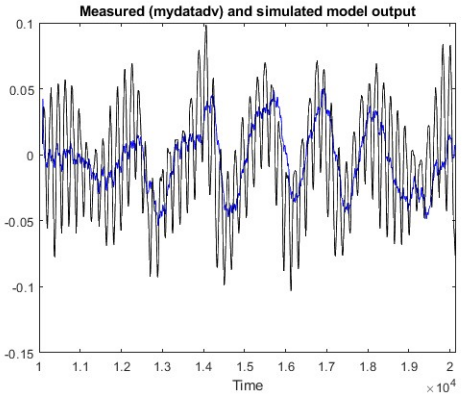


Figure 3. Scaled bioreactor temperature T data (black) and model (blue), time in minutes.

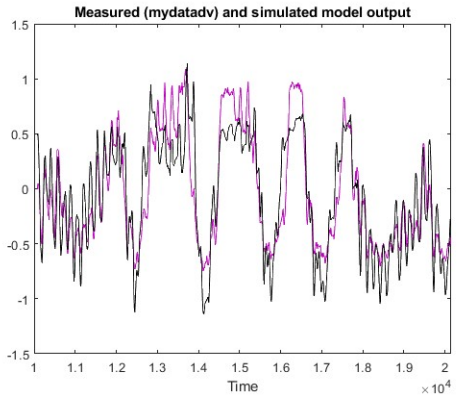


Figure 4. Scaled recycled sludge temperature out of heat exchanger T_{HX} data (black) and model (magenta), time in minutes.

3.4 Control

The control aim is to maximize biogas production F_{CH4} and minimize costs for pumping inlet sludge F_{in}, pumping recirculated sludge F_{HX} and heating sludge T_{HX}, while maintaining optimal temperature T in the bioreactor. Controlled variables are biogas produc-

tion F_{CH4}, bioreactor temperature T and recirculated sludge temperature T_{HX}. The manipulated variables are flowrate of sludge in F_{HX}, flowrate of recycled and heated sludge F_{HX}, and flowrate of hot water F_{HW}. The disturbance variables of the system are total suspended solids in TSS_{in}, inlet sludge temperature T_{in} and hot water temperature T_{HW1} into the heat exchanger Two control strategies were designed based on the existing control strategy at Veas WRRF and a recent master thesis work (Mukhtar, 2023). The first control strategy with three PID controllers is illustrated in Figure 5 and, the second control strategy with one model predictive controller is presented in Figure 6.

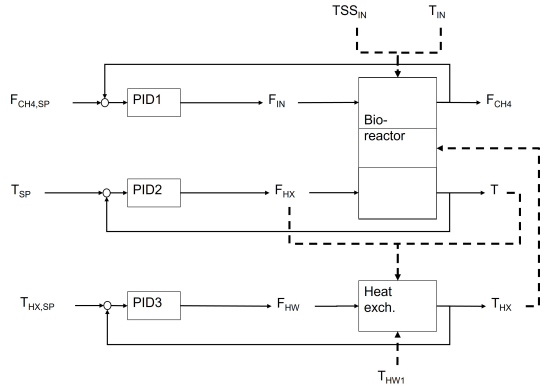


Figure 5. Suggested PID control strategy.

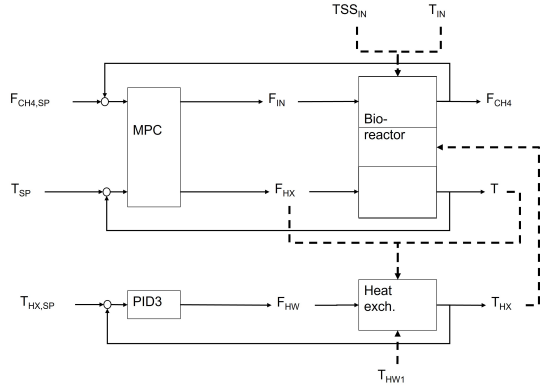


Figure 6. Suggested MPC strategy.

3.4.1 PID controllers

The PID controllers were parametrized using Skogestad tuning rules for first order system, the parameters are given in Table 6. The minimum and maximum limits for the PID controller outputs were minimum and maximum values from the scaled data of the manipulated variables.

Table 6. PID parameters.

Controller	Kc	Ti	tauc	min	max
PID1	0.5289	800	200	-2.03	1.45
PID2	-0.3998	19.7209	200	-0.14	0.19
PID3	0.0845	-	200	-12.00	15.48

3.4.2 Model Predictive Controller

The MPC controller tuning parameters are given in Table 7. The settling time was calculated as average between TFa and TFb. The MPC sampling time was chosen to keep the model horizon N, a ratio between settling time and sampling time under 120. Control horizon M was chosen between 1/3 and 1/2 of N. The prediction horizon was a sum of model horizon N and control horizon M. The output variables biogas flowrate F_{CH4} and bioreactor temperature T were weighed 10:1 (Q) to give more importance for the biogas production. Saturation limits for manipulated variables were the same as for PID controllers. Movements in the manipulated variables, F_{in}, F_{HX}, were restricted with Rd values. Tested Rd values included 1-1, 5-15, 10-60, 35-210, 50-300, 100-500. Through extensive simulation tests 35-210 gave lowest values on flowrate of inlet sludge F_{in}, reduced oscillations in both flowrates, and avoided flowrates to remain at saturation limits.

Table 7. MPC parameters.

MPC controller			
MPC sampling time	Ts	10	
Model horizon	N	$22 \cdot Ts$	
Control horizon	M	$11 \cdot Ts$	
Prediction horizon	P	$33 \cdot Ts$	
CV weights	Q	F_{CH4}	T
		2	1
MV saturation limits		F_{in}	F_{HX}
	min	-2.03	-0.14
	max	1.45	0.19
MV rate	Rd	F_{in}	F_{HX}
weights		35	210

3.4.3 Controller testing

The controllers were tested using the transfer functions as process model. The testing was performed using the plant data as disturbance variables (inlet sludge temperature T_{IN}, inlet sludge total suspended solids TSS_{IN} and hot water inlet temperature T_{HW1}). Setpoint for bioreactor temperature T and for recycled and heated sludge T_{HX} to follow a 240 sample moving

mean of the original data. As the aim is to improve biogas production, setpoint for biogas flowrate $FCH4$ was created as 240 sample moving mean multiplied by 1.00, 1.03, 1.05 and 1.10. Multiplication by 1.00 allows fair comparison between the real operation (orig.) and the proposed control strategies, whereas the increased biogas flowrate setpoint can show how much increase production will affect the manipulated variables. The integral of absolute error was calculated between setpoint and measurement for the controlled variables biogas flowrate $FCH4$ and bioreactor temperature T . Recycled and heated sludge T_{HX} is an intermittent variable between the bioreactor and heat exchanger, and therefore not added to the results table. Integral of absolute movement in the manipulated variables was calculated for inlet sludge flowrate F_{IN}, sludge recirculation rate F_{HX} and hot water flowrate F_{HW}.

The results for controlled variables in Table 8 show that biogas production $FCH4$ can be increased up to 10 % and variability (IAE) in controlled variables decreased with both PID and MPC strategies. The results for manipulated variables in Table 9 show that the MPC controller gives $92-99\%$ lower variability (IAMV) in the manipulated variables than the original control strategy, where as the PID strategy gives much higher variablity (IAMV) in the inlet sludge flowrate but $25-76\%$ lower for the other manipulated variables than the original control strategy. The integral of the scaled inlet sludge flowrate F_{in} and hot water flowrate F_{HW} have lowest values for the MPC controller without setpoint increase. When the biogas production setpoint is increased, naturally also inlet sludge flowrate F_{in} is increased.

Hence, the best control results for both controlled and manipulated variables are achieved with the MPC controller. The visual results for the scenario without setpoint increase are shown in Figures 7 - 11 .

Table 8. Control results CV

Control strategy	Int F_{CH4}	IAE F_{CH4}	IAE T	IAE T_{HX}
	$\cdot 10^6$	$\cdot 10^4$	$\cdot 10^2$	$\cdot 10^3$
Orig.	2.50	6.51	7.06	8.17
PID	2.50	1.85	0.92	1.62
PID 3%	2.57	1.90	0.92	1.62
PID 5%	2.62	1.94	0.92	1.62
PID 10%	2.75	2.04	0.93	1.62
MPC	2.50	1.89	2.86	1.61
MPC 3%	2.58	2.06	3.05	1.61
MPC 5%	2.62	2.21	3.10	1.61
MPC 10%	2.73	3.31	3.71	1.61

Table 9. Control results MV

Control strategy	Int F_{IN}	Int F_{HW}	IAMV F_{IN}	IAMV F_{HX}	IAMV F_{HW}
	$\cdot 10^3$	$\cdot 10^3$	$\cdot 10^3$	$\cdot 10^3$	$\cdot 10^3$
Orig.	-0.018	-0.003	5.85	1.06	185.7
PID	-1.626	1.21	19.45	0.80	43.95
PID 3%	4.642	2.50	20.35	0.80	43.93
PID 5%	8.816	2.61	20.95	0.80	43.94
PID 10%	19.26	2.92	22.59	0.79	43.94
MPC	-3.000	-1.32	0.46	0.040	1.92
MPC 3%	2.933	-1.60	0.45	0.039	1.91
MPC 5%	6.297	-2.92	0.45	0.037	1.92
MPC 10%	14.31	-5.21	0.31	0.025	1.93

4 Discussion and Summary

Development of energy-efficient control methods is crucial to reach the EU waste water directive target of energy-neutral of WRRF operation. Therefore, continuous efforts should be made to implement novel control technologies at municipal and industrial WRRFs. The work on modeling and control strategy development has been done in collaboration with Veas municipal WRRF in Norway. One of the main challenges on for control strategy development is availability of industrial measurements necessary to parametrise a state-of-the-art anaerobic digestion model. Hence, in this work a simplified approach with linear dynamic models was chosen. Three transfer function models were identified to model the biogas production in a bioreactor, temperature in the bioreactor and outlet temperature of the recirculation heat exchanger. The modeling results show sufficient fit $14\%-60\%$ to the industrial data.

Based on the transfer functions, two control strategies with PID and MPC controllers were designed. The control results show that both PID and MPC strategies decrease the variablity in the controlled and manipulated variables. MPC gave the best results, increasing biogas production up to 10 % and decreasing variability in controlled variables by $50-80\%$ and $92-99\%$ in manipulated variables. The answer to our research question is both PID and MPC control algorithms can optimize energy consumption and maximize biogas production. Best results can be achieved with MPC algorithm. Our results indicate that implementation of advanced control technologies can improve the energy-efficiency of biogas production at WRRFs.

We suggest future modeling work estimate the biogas recirculation rate, for example using a Kalman filter and research on modeling methods feasible for sequential operation. Other data-driven modeling methods could be tested for example using time-series

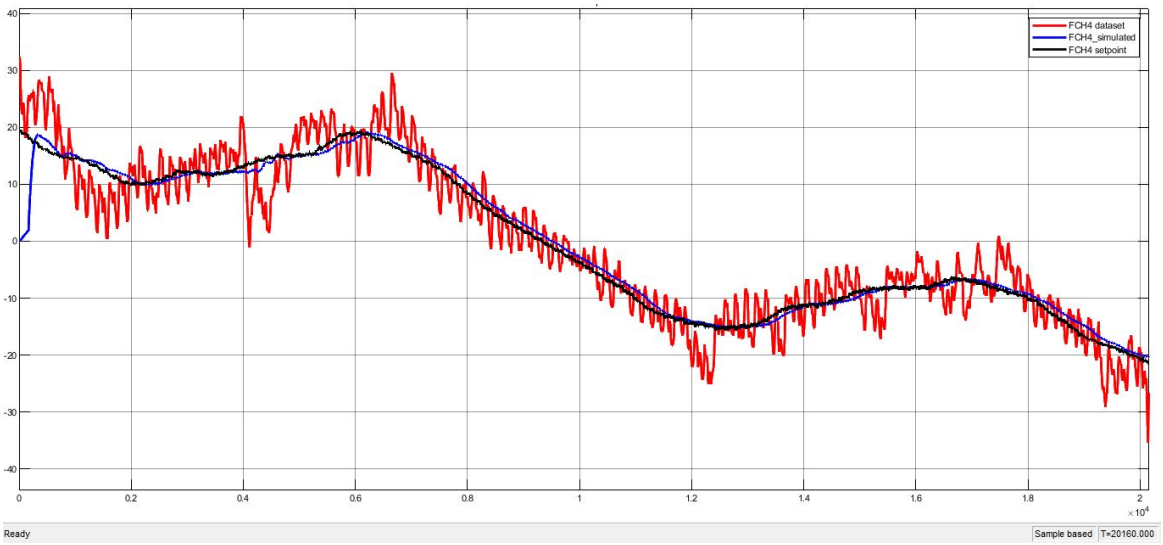


Figure 7. Biogas flowrate F_{CH4} original data (red), 24sample-moving average setpoint (black) and MPC simuled value (blue). Data is scaled, time in minutes.

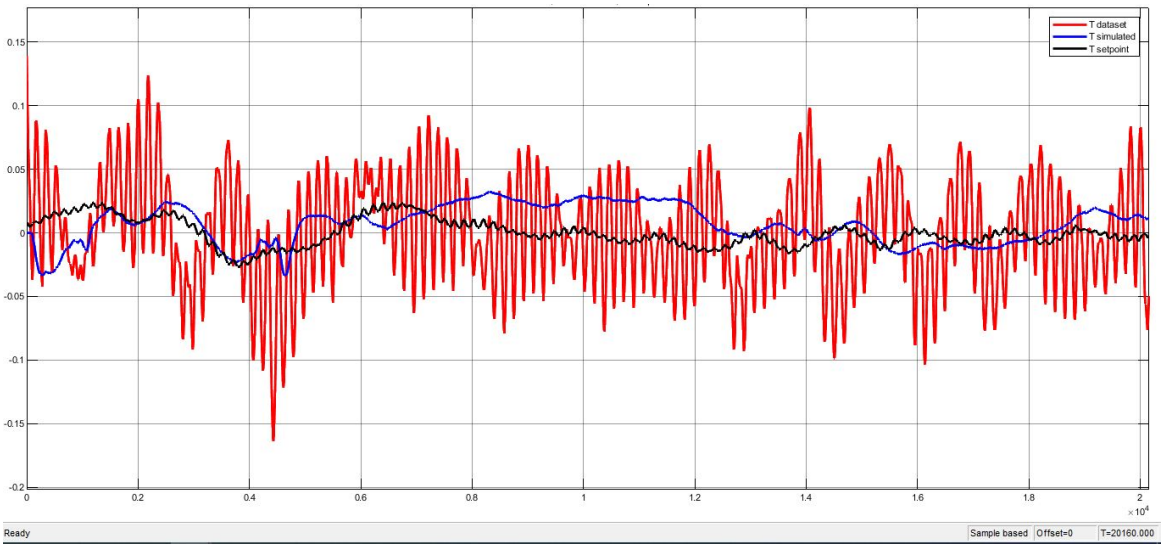


Figure 8. Bioreactor temperature T original data (red), 24sample-moving average setpoint (black) and MPC simuled value (blue). Data is scaled, time in minutes.

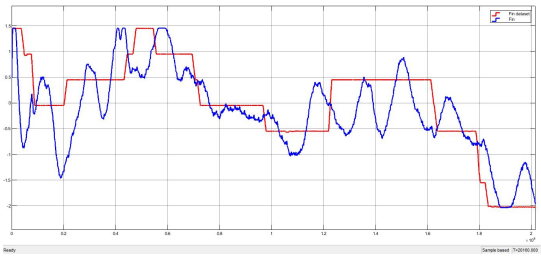


Figure 9. Inlet sludge flowrate F_{in} original data (red) and MPC simuled value (blue). Data is scaled, time in minutes.

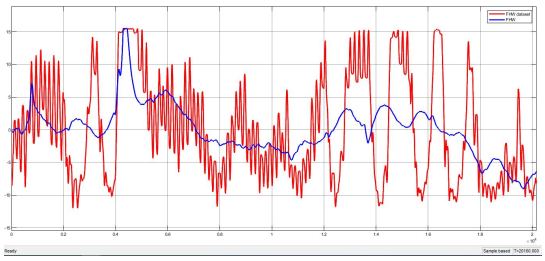


Figure 10. Hot water flowrate F_{HW} original data (red) and MPC simuled value (blue). Data is scaled, time in minutes.

models or Long-Short-Term-Memory networks could be tested. Future work on control should include predictive control algorithms that can account for sludge variations based on seasonality and weather progno-

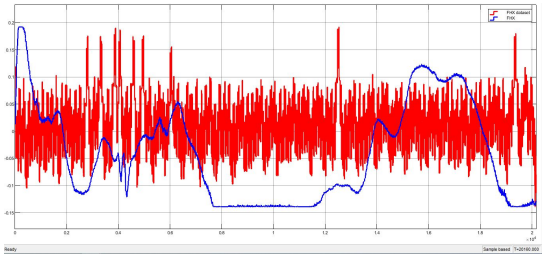


Figure 11. Sludge recirculation flowrate F_{HX} original data (red) and MPC simuled value (blue). Data is scaled, time in minutes.

sis. Adaptive control with AI approach could be explored.

5 Acknowledgements and Contributions

OsloMet is greatfully acknowledged for funding the MaxBiogas pre-project. Contributions:
Kristine Haualand, Kjell Rune Jonassen and Hilde Johansen have contributed to idea development and data collection from Veas wastewater resource recovery facility. Professor Tiina Komulainen is the principal investigator of the project. She has supervised modeling work and provided the initial ideas on modeling and control design and TF model structures. Master student Bilal Mukhtar has done initial work on identifying the transfer functions from Veas data, and setting together bioreactor model and control strategy. Research assistant Truls Ødegaard and Professor Komulainen have revised the biogas production model and control strategy based on Bilal Mukhtar's findings.
Head of studies Simen Antonsen has co-supervised Bilal Mukhtar's master project and done critical review of this article.

References

Attar, S., & Haugen, F. A. (2019). Dynamic model adaptation to an anaerobic digestion reactor of a water resource recovery facility. *Modeling, Identification and Control,*, *40*, 143–160. doi: doi:10.4173/mic.2019.3.2

Bergland, W. H., & Bakke, R. (2016). Modelling anaerobic digestion during temperature and load variations. *International Journal of Energy Production and Management*, *1*, 393–402. doi: https://doi.org/10.2495/EQ-V1-N4-393-402

EuropeanCommission. (2021a). *Circular economy action plan.* Retrieved from `https://ec.europa.eu/environment/strategy/circular-economy-action-plan_en`

EuropeanCommission. (2021b). *Eu strategy on energy system integration.* Retrieved from `https://energy.ec.europa.eu/topics/energy-system-integration/eu-strategy-energy-system-integration_en`

EuropeanCommission. (2022). *Proposal for a revised urban wastewater treatment directive.* Retrieved from `https://environment.ec.europa.eu/publications/proposal-revised-urban-wastewater-treatment-directive_en`

Komulainen, T. M., & Johansen, H. (2021). Possible concepts for digital twin simulator for wwtp. In *Proceedings of the first sims eurosim conference on modelling and simulation, sims eurosim 2021, and 62nd international conference of scandinavian simulation society, sims2021.* Virtual conference 21-23.9.2021: Linköping University Electronic Press. doi: https://doi.org/10.3384/ecp21185398

Ljung, L. (1999). *System identification - theory for the user* (2nd ed.). Prentice Hall. Retrieved from `https://onlinelibrary.wiley.com/doi/10.1002/047134608X.W1046`

Mukhtar, B. (2023). *Control of biogas production* . Oslo Metropolitan University.

NorskVann. (2017). *Nasjonal bærekraftstrategi for vannbransjen.* Retrieved from `https://norskvann.no/interessepolitikk/baerekraft-ma-prioriteres/`

Seborg, D. E., Thomas F. Edgar, D. A. M., & III, F. J. D. (2017). *Process control* (4th EMEA ed.). Wiley. Retrieved from `https://bcs.wiley.com/he-bcs/Books?action=index&itemId=111928595X&bcsId=10324`

Skogestad, S. (2003). Simple analytic rules for model reduction and pid controller tuning. *Journal of Process Control*, *13*, 291–309. doi: https://doi.org/10.1016/S0959-1524(02)00062-8

A Comparison of Strain Gauge Measurements and FEA for a Confined Channel Geometry Subjected to a Hydrogen-Air Mixture Explosion

Daniel Eckhoff [a, *], Magne Bratland [a], Mads Mowinckel [a]

[a] *Department of Process, Energy and Environmental Technology, University of South-Eastern Norway*
*danieeck@stud.ntnu.no

Abstract

Using finite element analysis for rapid dynamic loads without validation of the results can lead to major miscalculation, thus making it necessary to examine the accuracy of the software. The structural response from a hydrogen-air mixture explosion in a confined channel is investigated with experiments and numerical methods. The channel measures 1000 mm in length, with an inside diameter of 65 mm, and 15 mm thick transparent polycarbonate sidewalls. Hydrogen and air were released into the channel and ignited. Four Kistler transducers record the internal pressures. A biaxial HBM rosette strain gauge was bonded to the polycarbonate sidewall, used for recording strains during the explosion experiments, where von Mises stresses were calculated from these recordings. The channel was then idealized as a computer-aided design model in the engineering software SOLIDWORKS. By utilizing the pressure data from the experiments and creating a four-pointed loading curve, finite element analysis was applied for obtaining numerical von Mises stress results. Comparing the experimental and numerical results of von Mises stress show a variation of 4.9%.

Keywords: Structural Response, Finite Element Analysis, Strain Gauges, Explosions

1. Introduction

Hydrogen can be a key contributor in the transition to renewable energy, especially in the process and transportation sectors. As hydrogen is an efficient energy carrier with a zero emission of CO_2 during combustion, it will be an important resource for solving current environmental challenges. Despite this benefit, there is also a downside; hydrogen is a highly flammable substance with an associated risk of fire and explosion. An accidental explosion can have immense consequences such as economic losses, personnel injuries, or in worst case fatalities. With the following dangers of hydrogen, precautionary measures in engineering structures and designs transporting or storing hydrogen must be taken into consideration. This should be prioritized to limit the potential destructive outcome in case of an explosion.

A blast wave from a hydrogen explosion inside a confined space will be of a rapid phenomenon, combined with high pressures and energies [1]. As the explosion's subsequent waves will reflect inside confined spaces such as channels or tunnels, it can be challenging to numerically simulate the explosion and the structural response simultaneously as a coupled occurrence. The normally applied approach is using decoupled numerical simulation, where the structure is modeled in a computer-aided design (CAD) software and furthermore simulated with a less complex explosion incident using the finite element method (FEM). The loading scenario in these simulations are often simplified to a curve in a triangular shape for pressures, forces, or velocities, where this data can be obtained through experiments, distinct numerical simulation or using empirical formulas [2-4]. However, this process requires certain assumptions for both the CAD model and the loading curve. It is hence necessary to verify the accuracy of the results from finite element analysis (FEA) software. This is especially important when performing FEA with complex loading situations such as explosions or other rapid dynamic loads.

In this paper, the occurring von Mises stresses on a 1-meter-long alloy steel channel exposed to a hydrogen-air mixture explosion is analyzed, with physical experiments using biaxial rosette strain gauges, and using decoupled numerical simulation. It is also beneficial to use von Mises stresses as a reference for comparison, as the strain gauge can measure biaxial stresses under complex explosive loads. Furthermore, the aim of this study is to verify

if the simulation results from the commercial computer-aided engineering (CAE) software SOLIDWORKS[1] will be similar to the strain gauge measurements, and if the software is suitable to use for these kind of loading circumstances. This information will be helpful for engineers designing structures which should withstand explosive loads.

The experiments were conducted in collaboration with another research project at the University of South-Eastern Norway, where the physics of the explosion and the deflagration-to-detonation transition (DDT) was investigated [5]. This paper will focus on the structural behavior of the experimental rig.

2. Experimental setup

The experimental rig consisted of an alloy steel channel with an inside length of 1000 mm and a cross section of 116x65 mm. The channel was closed at the left end, and open at the right end. On its sidewalls, it was fitted with 15 mm thick transparent polycarbonate sheets fastened with M8 bolt connections, tightened to 15 Nm. Two steel plates were mounted at the right end for structural stability. Recording of pressure data were done by four Kistler pressure transducers (P1-P4) with a sampling rate of 100 kHz, mounted to the channel with a spacing of 250, 450, 650 and 850 mm relative to the left side of the channel, see Fig. 1. The center of the channel was filled with 40 cylinders, with an intent to create a turbulence for the gas and incite a DDT.

For strain gauge measurements, an HBM 3/120 RY81 Rosette with three measuring grids and a resistance of 120 Ω was used. The gauge factor was 2.03 for all measuring grids, with a transverse sensitivity of 0.2 % for gitter A and C, and a 0% sensitivity for gitter B. HBM states that the measurement uncertainty for stress measurements can be up to 6% [6]. The strain gauge was wired in a half bridge to three channels to an HBM Spyder 8 data acquisition module (DAQ), with a sampling rate of 4.8 kHz. The DAQ module was connected to a computer running the catman[2] measuring software. Catman was set-up to calculate von Mises stresses. The strain gauge was bonded to the backside polycarbonate sheet, see Fig. 2.

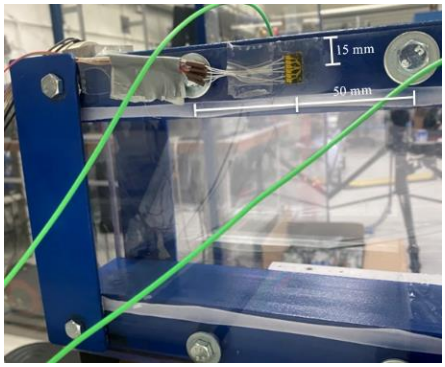


Figure 2: Strain gauge location

By initiating a spark to the mixture of hydrogen and air in the channel, a following explosion would occur.

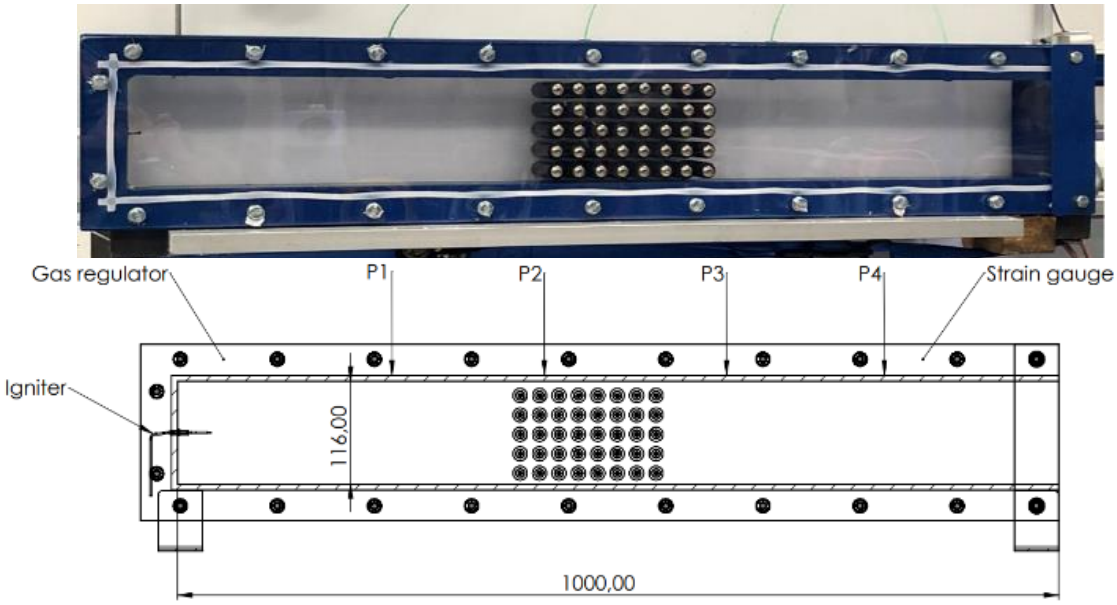


Figure 1: Photo and schematic of the channel

[1] SOLIDWORKS by Dassault Systèmes
[2] catman Data Acquisition Software by HBM

The transducers recorded the achieved explosion pressures, which would later be used in the FEA. A more in-depth explanation of the initiating explosion procedure is given by Henriksen et al. [5]. In total, four experiments were conducted.

Furthermore, as the pressures recorded from transducer P4 are the most interesting (since the strain gauge is bonded in this area), this pressure data will be used in the FEA analysis. A table showing the P4 transducer peak pressures can be seen in Tab. 1. The pressure curve from Exp. 2 for P4, which was the highest achieved pressure, is shown in in Fig. 3.

Tab 1: Peak pressures from P4

	Peak pressure recorded by transducer P4 [MPa]
Exp. 1	0.42
Exp. 2	1.21
Exp. 3	0.58
Exp. 4	1.19

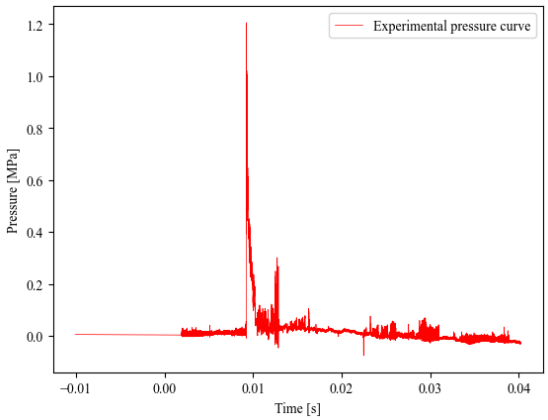


Figure 3: Pressure curve from Exp. 2 P4 transducer

3. Numerical setup

The experimental rig was modeled in SOLIDWORKS and simplified to a 260 mm long symmetric model relative to the open end of the channel. For meshing the model, a blended curvature-based mesh with a minimum element size of 3 mm and maximum element size of 20 mm was used. The model consisted of a total of 24739 four-nodal tetrahedral elements and 41289 nodes. 98% of the mesh had an aspect ratio lower than 3.

Using the SOLIDWORKS Connection feature, six M8 bolts with a pretension torque of 15 Nm were added to the model. Contact feature was used to simulate the physical contact of the channel, steel plate and the polycarbonate sheet. The model and the mesh can be seen in Fig. 4. As the highest explosion pressures were recorded in Exp. 2 (see Tab. 1), the simulations were based of this data.

Furthermore, it was used symmetry conditions to enforce boundary conditions.

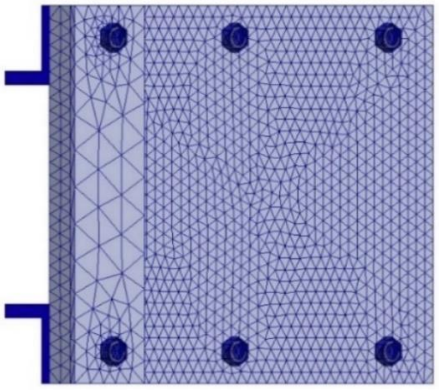


Figure 4: Meshed FEA model

By utilizing the peak pressures recorded by P4 during Exp. 2, the loading pressure curve was simplified to a four-pointed triangular-shaped curve relative to time, see Fig. 5. The pressure was uniformly placed on all inside surfaces of the model. Furthermore, the simulation was run as a non-linear dynamic study with 59 steps with an initial time increment of 0.001 sec., starting at 0 sec. and ending at 0.012 sec. The simulation study was conducted up to a time period of 0.012 seconds due to the fact that the explosive blast pressures from Exp. 2 diminish to nearly 0 MPa at the end this timeframe.

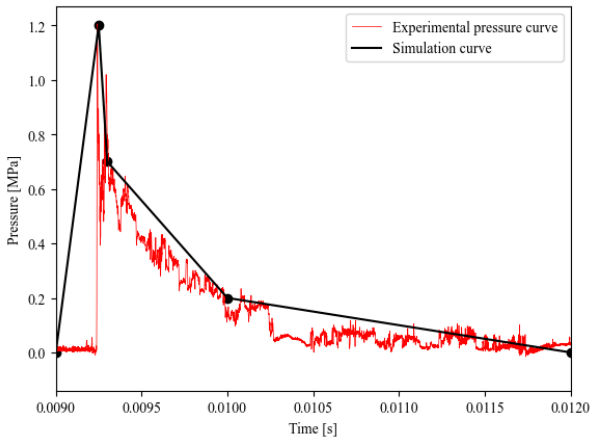


Figure 5: Simulation loading curve compared with Exp. 2 P4 pressure curve.

4. Results

This section is divided into three subsections: experimental results, numerical results and comparison of the results.

4.1. Experimental results

The maximum achieved internal explosion pressure from P4 is shown in Tab. 1, as this is the area of the strain gauge location. The measured von Mises stresses results from the experiments can be seen in Fig. 6.

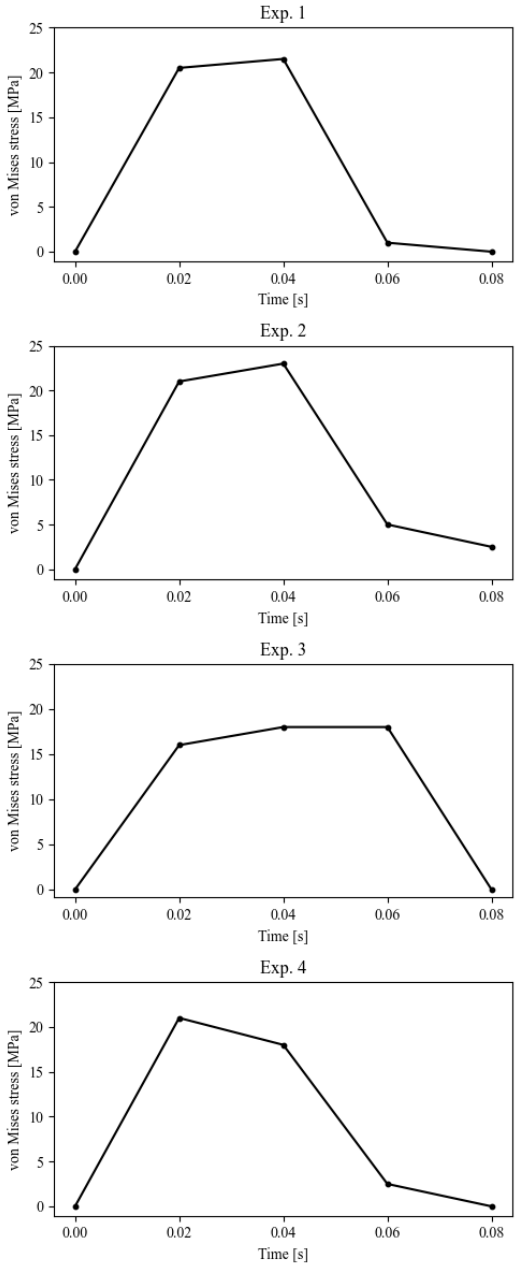


Figure 6: Strain gauge measurement results

The values seem to be similar in their maximum von Mises stress, alternating between 19 MPa and 22 MPa. The DAQ recorded five data points of stress for the experiments. This can be seen in the x-axis presenting time, starting at 0 sec. to 0.08 sec. The data points have a spacing of 0.02 sec.

Data such as hydrogen flow, ignition timing/position or if DDT occurred is beyond the scope of this paper, and thus not emphasized, nor presented.

4.2. Numerical results

Numerical results from the non-linear dynamic SOLIDWORKS simulation achieved a peak stress of 22.2 MPa in the same area as the strain gauge was located. This happened in simulation step 49 at 0.009 sec simulated time., which was the time of the highest loading pressure. The scaling of time in the dynamic response is different from the von Mises experimental results, as the loading pressure curve and simulation are based on data from the pressure transducer P4, see Fig. 7. In Fig. 8, the FEA stress contour plot of the polycarbonate plate model is presented.

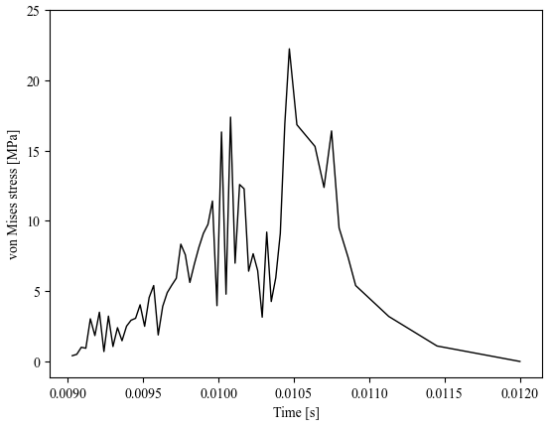


Figure 7: Structural simulation response

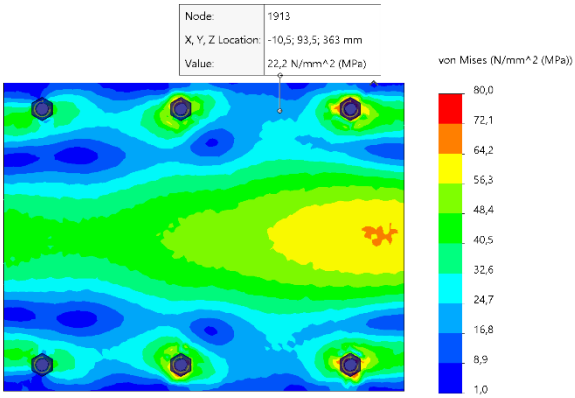


Figure 8: Stress contour plot from FEA

4.3. Comparison of experimental and numerical results

The achieved von Mises stresses are compared as a result. The maximum recorded von Mises stress in Exp. 2 was 21.85 MPa compared to the FEA simulation with 22.2 MPa, which is a difference of 4.9%. See Fig. 9 for a graphical representation of the comparison of achieved stresses, and Fig. 10 for a highlighted area of interest in stress comparison.

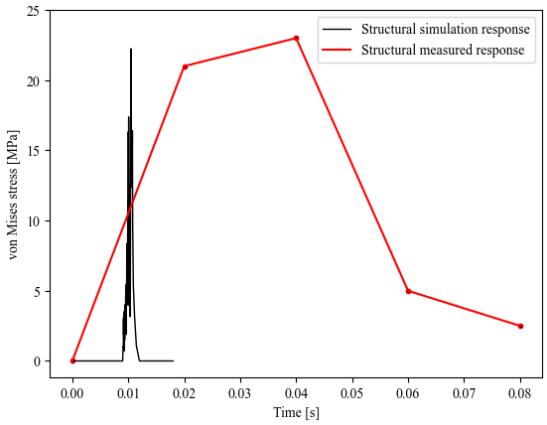


Figure 9: Numerical response compared to experimental for Exp. 2

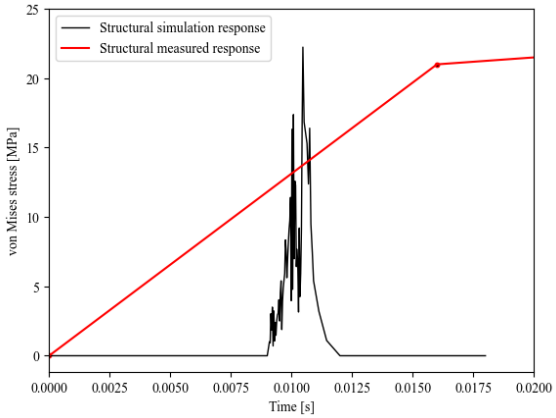


Figure 10: Highlighted area of interest in stress comparison of Fig. 9.

5. Discussion

The purpose of this paper was to investigate and compare von Mises stresses from experiments and FEA using SOLIDWORKS Simulation. For the experiments, biaxial strain gauges connected to an HBM DAQ recorded at least two data points at the polycarbonate sheet's stress peak. The numerical simulation was based on the decoupling method. However, instead of using a standard triangular shape for the loading curve, a four-pointed curve (Fig. 5) was used in the simulation. The simulation was run as a non-linear dynamic study. The comparison of the experimental and numerical results shows similarities, as both reached a von Mises stress peak at 22 MPa (± 0.2 MPa).

Since the simulation was based on the recorded internal pressures from the experiments, so was the scaling of time. This resulted in a time scale from 0 sec. to 0.012 sec., set against the strain gauge measurements which scaled from 0 sec. to 0.08 sec. In the simulation result in Fig. 7, the stresses on the polycarbonate reached a maximum stress peak at 0.0107 seconds, with a decreasing value after the peak, ending at 0.012 sec. (also the time the simulation ended). The stresses from the strain gauge measurements did not start to decrease until 0.04 sec for 3 of the 4 experiments. This could possibly be due to the structure's inertia [7] or the DAQ not being able to record smaller stress alternations. During the experiments, a build-up of the loading explosion pressure occurred over a rapid time interval. However, in the simulation, the maximum peak pressure of 1.2 MPa was applied to the structure over a slower time interval of the simulation study. This suggests the possibility for a difference in the stress alternations.

The sampling rate of the DAQ used for strain gauge measurements was low, running at 4.8 kHz versus the pressure transducers running at 100 kHz. This could result in the DAQ being uncapable of recording stresses that could potentially be higher than 21.85 MPa, or stresses occurring between the 0.02 sec. sample intervals. In Fig. 9/10, it is shown that the simulated stresses achieve close to 22 MPa almost immediately, unlike the measured stresses from Exp. 2 which does not achieve any stress peaks prior to 0.02 sec. However, none of the experiments reached a higher total von Mises stress than 21.85 MPa (Exp. 2), while the experiments also were consistent in the measured stresses. The possibility for stress peaks reaching higher than 21.85 MPa between the 0.02 sec. intervals for a total of 4 experiments therefore seems low, meaning that the maximum stress measurements appear to be reasonable.

Limitations of the measuring equipment do clearly give an inaccuracy of the time scaling for comparison of the results. Nevertheless, as this paper focused on analyzing the maximum occurring stresses in the polycarbonate sheet sidewall, the time scale correctness is not especially relevant for this study. In further work, a higher sampler rate DAQ should be used for the possibility of having an exact time scale and to reduce the potential for measurement uncertainty.

6. Conclusion

This study demonstrates the use of FEA with SOLIDWORKS for a decoupled numerical simulation, and how the results compare to physical experiments using strain gauges to obtain von Mises stresses. The results show a difference of 4.9% of the maximum achieved stresses. This information and procedure can be helpful for design engineers constructing structures for withstanding explosive or other rapid dynamic loads.

CRediT authorship contribution statement

Daniel Eckhoff: Conceptualization, Methodology, Formal analysis, Investigation, Writing - Original Draft, Writing - Review & Editing.

Magne Bratland: Conceptualization, Writing - Review & Editing, Supervision.

Mads Mowinckel: Conceptualization, Methodology

Acknowledgments

The authors would like to thank the research group Telemark Process Safety, Combustion and Explosion Laboratory, and especially Associate Professor Mathias Henriksen, for aid and assistance regarding the experiments.

References

[1] G. L. Blanc, M. Adoum, and V. Lapoujade, “External blast load on structures – Empirical approach,” p. 12.

[2] Y. Lu and Z. Wang, “Characterization of structural effects from above-ground explosion using coupled numerical simulation,” Comput. Struct., Aug. 2006.

[3] J. Zhu, L. Yashi, S. Wu, Z. Ru, and L. Ren, “Decoupled explosion in an underground opening and dynamic responses of surrounding rock masses and structures and induced ground motions: A FEM-DEM numerical study,” Tunn. Undergr. Space Technol., vol. 82, pp. 442–454, Dec. 2018, doi: 10.1016/j.tust.2018.08.057.

[4] Y. Du, L. Ma, J. Zheng, F. Zhang, and A. Zhang, “Coupled simulation of explosion-driven fracture of cylindrical shell using SPH-FEM method,” Int. J. Press. Vessels Pip., vol. 139–140, pp. 28–35, Mar. 2016, doi: 10.1016/j.ijpvp.2016.03.001.

[5] M. Henriksen, K. Vågsæther, and D. Bjerketvedt, “Deflagration-to Detonation Transition of Hydrogen-Air Mixture in a Highly Congested, Open-ended Channel,” Unpubl. Manuscr. Fac. Technol. Nat. Sci. Marit. Sci. USN Porsgrunn, Mar. 2022.

[6] “Measurement Uncertainty | Experimental Stress Analysis,” HBM, Oct. 07, 2020. https://www.hbm.com/tw/6021/measurement-uncertainty-experimental-stress-analysis/#:~:text=The%20uncertainty%20of%20the%20gauge (accessed Aug. 17, 2023).

[7] V. V. Molkov, “Turbulence generated during vented gaseous defla- grations and scaling issue in explosion protection,” no. 148, p. 14, 2001.

A deep learning approach for fault diagnosis of hydrogen fueled micro gas turbines

Muhammad Baqir Hashmi [a, *], Mohammad Mansouri [a, b], Amare Desalegn Fentaye[c], Shazaib Ahsan[d]

[a]*Department of Energy and Petroleum Engineering, University of Stavanger, 4036, Stavanger, Norway,*

[b]*NORCE Norwegian Research Centre, 4021, Stavanger, Norway,*

[c]*School of Business, Society and Engineering, Mälardalen University, 883, SE-72123 Västerås, Sweden,*

[c]*Department of Mechanical Engineering, University of Manitoba, Winnipeg, Manitoba, R3T 5V6, Canada*

Corresponding author email : muhammad.b.hashmi@uis.no

Abstract

Hydrogen fueled gas turbines are susceptible to rigorous health degradation in form of corrosion and erosion in the turbine section of a retrofitted gas turbine due to drastically different thermophysical properties of flue gas stemming from hydrogen combustion. In this context fault diagnosis of hydrogen fueled gas turbines becomes indispensable. To authors knowledge, there is a scarcity of fault diagnosis studies for retrofitted gas turbines considering hydrogen as a potential fuel. The present study, however, develops an artificial neural network (ANN) based fault diagnosis model using MATLAB environment. Prior to fault detection, isolation and identification modules, physics-based performance data of 100 kW micro gas turbine (MGT) was synthesized using GasTurb tool. ANN based classification algorithm showed a 99.4% classification accuracy of fault detection and isolation. Moreover, the feedforward neural network-based regression algorithm showed quite good training, testing and validation accuracies in terms of root mean square error (RMSE). The study revealed that presence of hydrogen induced corrosion fault (both as single corrosion fault or as simultaneous fouling and corrosion) led to false alarms thereby prompting other wrong faults during fault detection and isolation modules. Additionally, performance of fault identification module for hydrogen fuel scenario was found to be marginally lower than that of natural gas case due to assuming small magnitudes of faults arising from hydrogen induced corrosion.

1. Introduction

The power sector was responsible for ~38% of the global carbon dioxide emissions in 2021. Natural gas (NG) reportedly contributed to ~22% of the electric power generation globally in 2021 (EDGAR/JRC. 2022). By far, gas turbines are mainly burning NG for power generation resulting in greenhouse gas (GHG) emissions and climate change. Therefore, decarbonization of gas turbines becomes indispensable to meet global energy transition mandate. In this context, the gas turbine industry aims for 100% carbon neutral gas fired power generation using low carbon fuels such as hydrogen by 2030 (TURBINE).

Nevertheless, utilization of hydrogen in gas turbines raises several technological and reliability challenges due to radically different thermophysical properties of hydrogen as compared to NG. For instance, hydrogen can potentially lead to flashback and thermoacoustic instabilities in lean premixed dry low emissions / dry low NO_x (DLE/DLN) burners. Flashback can damage the upstream components of the burner. It is worth noticing that available DLN technologies are currently capable of burning up to 60% hydrogen (Noble et al. 2021). The utilization of 100% hydrogen needs reconfiguration of the gas turbine with a new hydrogen compliant burner and modified fuel system. However, high hydrogen utilization produces enhanced steam content in the combustion flue gas that in turn is responsible for high heat transfer to the metal parts, higher thermal conductivity, aggravated oxidation corrosion, increased creep and thermal fatigue damages of hot gas path components (Gazzani et al. 2014). In this regard, hydrogen fueled gas turbines are susceptible to more health degradation caused by already mentioned problems. Especially, the retrofitted gas turbines, in which solely burner is replaced with hydrogen compliant burner keeping the existing turbomachinery, have more propensity of health degradation. Therefore, intelligent fault diagnosis, prognosis and health monitoring is of crucial importance for enhanced reliability and availability of hydrogen fueled gas turbines.

Normally, as the operating hours of gas turbines increase, performance and health degrade due to

various components faults (Marinai, Probert, and Singh 2004). To carry out effective maintenance actions, timely fault detection and identification play a key role in assuring reliability of the engines. Fault d diagnosis has been used over the years for industrial and aero gas turbines. It started with Urban's rudimentary concept of linear gas path analysis (GPA) (Urban 1975; Urban and Volponi 1992). Nowadays, gas turbine gas path diagnostic is typically carried out via three methodologies, i.e., model based, data driven and hybrid approaches (Fentaye et al. 2019).

For micro gas turbines, there are a few studies relevant to performance-based fault diagnosis that considered radial compressor fouling, turbine erosion and recuperator degradation phenomena, all considering NG fueled scenarios. Gomes et al., (Gomes et al. 2006) reported that presence of the recuperator in MGT increases the sensitivity of engine to compressor fouling and turbine erosion especially in variable speed operating mode. Hence, they conducted a comparative study of several single and multiple faults i.e., fouling, erosion, foreign object damage (FOD) and recuperator deterioration. The study adopted a model-based approach namely NLGPA technique using Pythia and Turbomatch tools for fault diagnosis. Another study conducted by Yoon et al., (Yoon et al. 2008) employed neural networks for prediction of degraded performance of a 30 kW MGT. Various kinds of single and multiple faults in compressor, turbine and recuperator were included in the study. The approach was found to be predicting the results with much accuracy even if some measurements data were missing.

Talebi and Tousi (Talebi and Tousi 2017) attributed compressor fouling as one of the majorly occurring faults in the MGT engine and hence they investigated the effect of blade surface roughness on the performance degradation of radial turbomachinery in a 477 kW MGT. The study revealed that combustor inlet temperature and turbine outlet temperature were more sensitive to blade surface roughness because these measurements showed increased values than the allowable limits. However, compressor discharge temperature was found to be less sensitive to the roughness. In a similar study, Bauwens (Bauwens 2015) also asserted that compressor fouling was a highly likely occurring fault in a 3 kW MTT MGT because of the possibility of oil ingestion in the compressor originating from de-aerating oil sump. Talebi et al., (Talebi et al. 2022) utilized artificial neural network (ANN) for fault detection and isolation of a 100 kW MGT considering the measurements uncertainties at different part load settings.

After an in-depth literature study, it seems that corrosion study of MGTs fueled by NG is scarce. It was also found that hydrogen induced corrosion in hot gas path components of both larger gas turbines and MGTs had not been investigated before based on authors' best knowledge. These research gaps paved the way for developing a fault detection, isolation, and identification model for a 100 kW MGT running with pure hydrogen fuel.

The present study incorporates a thermodynamic model using the commercial tool GasTurb 14 for generating a validated design point and off design performance data. Data preprocessing was implicated for adding noise and correcting the data for ambient condition variations. Subsequently, the data was fed to classification and regression learner tools in MATLAB version 2022a for fault detection and diagnosis purposes using neural network approach.

2. Methodology

The overall methodology of the entire study consists of 7-steps as illustrated in Fig. 1.

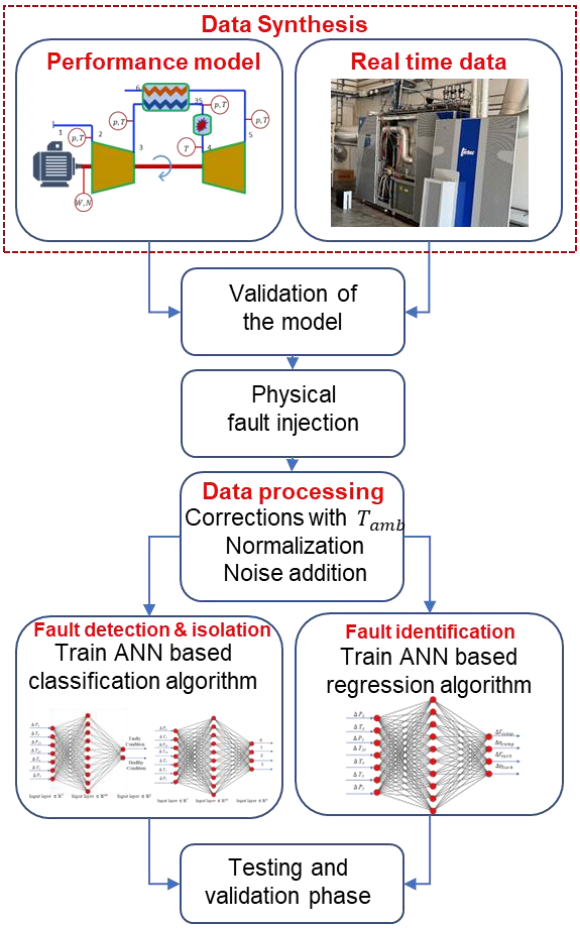


Figure 1: Detailed flow path of the methodology

The process includes developing a physics-based performance model, validating the model with real time MGT data, implanting the physical faults using health parameters i.e., flow capacity and efficiency, processing of synthesized performance data, fault detection and isolation (FDI), fault identification,

and finally testing and validation of the algorithms. Data processing is further segregated into correcting the data against ambient conditions, finding measurement deltas of the signals, and noise addition. Subsequently, the corrected-measured-noisy data of the signals are fed to ANN based classification and regression algorithms for developing a holistic fault diagnosis model. Different steps of the fault diagnosis process are illustrated in Fig. 1. The details of these steps are described in the following sub-sections.

2.1. Baseline performance model

A thermodynamic performance model of a 100 kW MGT was initially developed using commercial software tool GasTurb (Kurzke 2012) for physics-based data generation. The schematic of the MGT with sensor measurement points at various gas path stations is illustrated in Fig. 2.

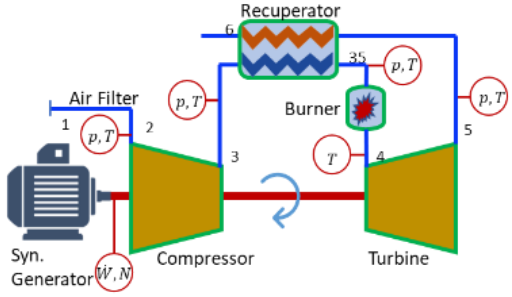


Figure 2. Schematic of a 100 kW MGT

The stations numbers have been identified at their respective positions such as 2 as compressor intake, 3 as compressor exhaust, 35 as recuperator cold side exhaust, 4 as turbine inlet, 5 as turbine exhaust and 6 as recuperator hot side exhaust. The design point calculations were optimized using random search algorithm to assure the accuracy of the baseline model. The off-design performance was calibrated with experimental data for accurate validation purposes. The experiments were conducted at different power settings varying from 50 to 100 kW with a step change of 10 kW. The ambient temperatures were noticed to be varying between 281.15 to 287.15 K during the entire test campaign. The real time data were taken by installing different pressure and temperature sensors in form of probes. To measure gas path conditions at intake of compressor, five pressure and four temperature measuring sensors were installed. Similarly, at compressor exit, three pressure and three temperature measuring sensors were installed 120° apart at circumferential positions to measure the average values at the flow field. Additionally, combustor head was also encompassed with pressure and temperature sensors to measure the conditions of intake air preheated by the recuperator. The instruments used for measuring pressure at different points were Kiel probes installed ±35° apart. Pressure scanners were adopted to scan the pressure with an accuracy of 0.05 of full-scale output. Similarly for temperature measurements K-type thermocouple with an accuracy of ±1 K were installed, and data acquisition (DAQ) device was utilized to get the measured data. Subsequently, the pressure scanner and DAQ were connected to a computer in parallel mode via two ports which led to data visualization through LabView software. The validated design point data is listed in Tab. 1.

Table 1: Design point validation after optimization

Parameter	OEM data (TURBEC 2017)	Present study	% Error
Power output [kW]	100(±3)	100.1	0.09
Electrical efficiency [%]	30 (±1)	29.99	0.03
Pressure ratio [-]	4.5	4.5	0
Exhaust mass flow	0.8	0.799	0.12
Exhaust gas temperature [K]	543	556.83	2.5

The validated off-design data at different part load power settings for different measurement points are illustrated in Fig. 3 and Fig. 4. Firstly, the engine was simulated by assuming NG as a working fuel that basically established a baseline for further model development. Subsequently, hydrogen was utilized as a fuel that was the prime objective of the study. Both simulation scenarios, i.e., NG and hydrogen fuel were further utilized for appending measurement uncertainties along with ambient temperature corrections. Finally, this data was made ready for classification learning and artificial neural network (ANN) to carry out fault detection, isolation, and identification. However, P_3, T_3, P_{35}, T_{35}, T_4, and T_5, measurement signals were identified as the most significant parameters for fault diagnosis purpose based on their deviating fault signatures.

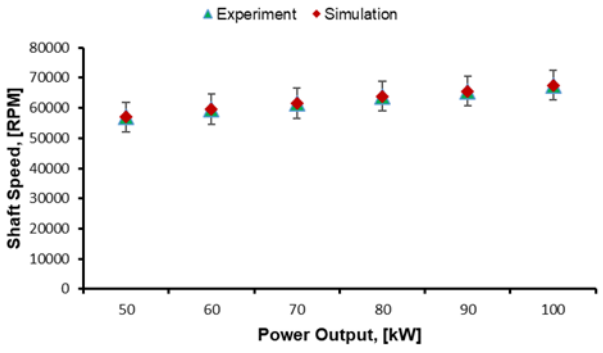


Figure 3. Shaft speed at different power settings

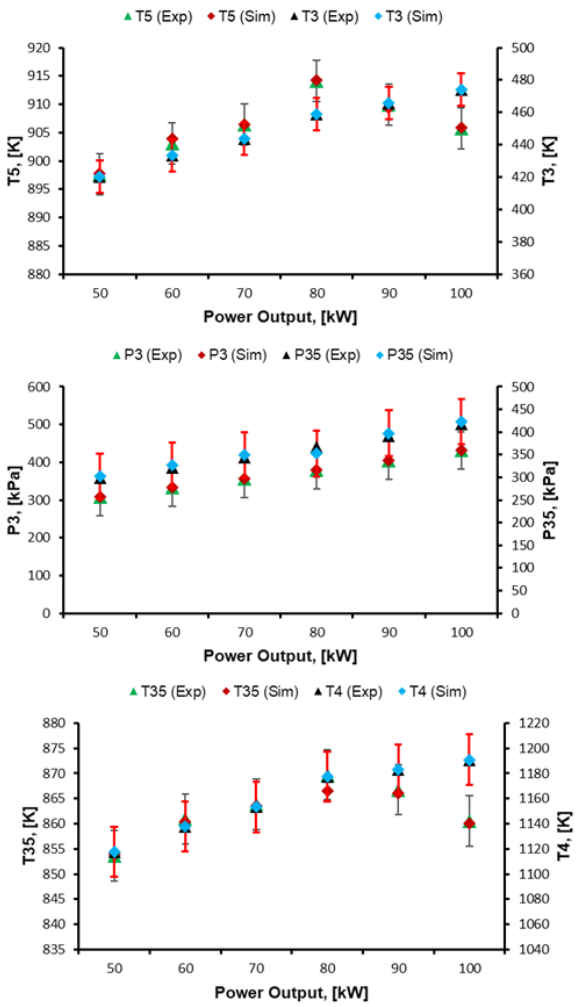


Figure 4: Different pressure and temperatures at various power settings

2.2. Component level degradation

The physical faults such as fouling, corrosion, erosion, and FOD occurring in different components of gas turbine systems lead to variations in engine health parameters or independent parameters i.e., flow capacity and isentropic efficiencies. These health parameters in turn stimulate deviations in engine measurements or dependent parameters such as pressure, temperature, fuel flow and shaft speed. The present study employs the nonlinear GPA (NLGPA) approach for gas path diagnostics of the MGT because of its added advantage over LGPA in terms of accuracy. This is a model-based diagnosis approach that employs a thermodynamic relationship between dependent and independent parameters addressing the nonlinearity of the gas turbine engine. The correlations are as follows,

$$\Delta\vec{Z} = H.\Delta\vec{X} \quad (1)$$

$\Delta\vec{Z}$ is a vector of measurement deviations of a degraded engine condition from clean condition. The clean condition is normally assumed as the healthy condition of the engine at design point. $\Delta\vec{X}$ expresses the health parameters and H represents the influence coefficient matrix (ICM) that develops a correlation between $\Delta\vec{Z}$, and $\Delta\vec{X}$. The further details can be found in the existing literature (Fentaye et al. 2019; Tahan et al. 2017).

The current study encompassed two kinds of component faults i.e., compressor fouling and turbine corrosion. The reason for choosing fouling is due mainly to a higher probability of occurring fouling in recuperated MGT as evidenced by the literature (Gomes et al. 2006; Bauwens 2015). Turbine corrosion was selected because hydrogen fuel leads to an enhanced steam content that can cause higher corrosion and heat transfer rates as compared to a NG fueled gas turbine (Oluyede and Phillips 2007). These hydrogen specific attributes can further lead to aggravated creep and material degradation in hot gas path components and hence to a reduced lifetime of the gas turbine. The quantification of components' physical faults is carried out by developing scaling factors of the health parameters (Flow capacity:Γ, Efficiency: η) as follows,

$$\Gamma_{deg} = \Gamma_{clean}\,(1 + \Delta\Gamma/100\,) \quad (2)$$

$$\eta_{deg} = \eta_{clean}\,(1 + \Delta\eta/100) \quad (3)$$

In the above-mentioned equations, subscript "deg" represents degraded component condition while "$clean$" represents engine's clean or healthy engine condition. Whereas health parameters are represented in their respective symbol as follows, (Flow capacity:Γ, Efficiency: η). However, Δ denotes the change in health parameters. To develop fault diagnosis models for gas turbines, a variety of fault magnitudes have been assumed by the literature that show a relative change of flow capacity and isentropic efficiency from the clean condition in form of scaling or correction factors. Tab. 2 lists the values of compressor and turbine degradation magnitudes with the respective ratios. It is worth mentioning that the fault magnitude of the fouling has been assumed similar for both fuel scenarios while fault magnitude of the hydrogen fuel scenarios has been assumed higher as compared to NG scenarios. The assumption for steam induced corrosion has been borrowed from a study by Zwebek and Pilidis (Zwebek and Pilidis 2004; Zwebek and Pilidis 2003), that was conducted for fault diagnosis of the steam turbine. The reason lies in the fact that steam induced corrosion led by hydrogen fuel behaves similar for both steam turbine and gas turbine.

Table 2: Quantification of various physical faults

Fault	FC (X)	Eff. (Y)	Ratios (X: Y)	Ranges	Ref.
Natural gas case					

CF	$\Gamma_c \downarrow$	$\eta_t \downarrow$	~3:1	[0, -7.5] [0, -2.5]	(Qingcai et al. 2016; Mohammadi and Montazeri-Gh 2014)
TC	$\Gamma_t \uparrow$	$\eta_t \downarrow$	~2:1	[0, 4] [0, -2]	(Escher 1995)
Hydrogen case					
CF	$\Gamma_c \downarrow$	$\eta_t \downarrow$	~3:1	[0, -7.5] [0, -2.5]	(Qingcai et al. 2016; Mohammadi and Montazeri-Gh 2014)
TC	$\Gamma_t \uparrow$	$\eta_t \downarrow$	~2:1	[0, 5] [0, -2.5]	(Zwebek and Pilidis 2003; Gomes et al. 2006)

FC: Flow capacity, Eff.: Isentropic efficiency, CF: Compressor fouling, TC: Turbine corrosion

2.3. Fault diagnosis

The diagnosis of the gas turbines is normally performed into three steps i.e., fault detection, fault isolation and finally fault identification. Fault detection provides information about the presence of any imminent physical abnormality in the system. Fault isolation helps in determining the exact type and location of the fault. Fault identification determines the severity magnitude of the any physical fault. The present study incorporated all these steps involved in the diagnosis.

2.3.1. Data processing

Prior to fault diagnosis of the MGT the data generated from the performance model went through preprocessing phase. During preprocessing, the data was first segregated on fuel basis i.e., NG and hydrogen. Subsequently, a fault wise segregation (i.e., compressor fouling, turbine corrosion, and simultaneous compressor fouling and turbine corrosion) was carried out. Temperature corrections was also considered to avoid the influence of the ambient temperature variations on the measurement signals, as follows,

$$\theta = \frac{T_{measured}}{288.15K} \quad (4)$$

The θ, in above equation is the correction factor of the measured temperature ($T_{measured}$) with respect to ambient temperature that is 288.15 K.

Measurement deviations of degraded conditions from clean condition of each signal i.e., P_3, T_3, P_{35}, T_{35}, T_4, T_5, and P_5 were estimated using the following relation,

$$\Delta\vec{Z} = \frac{(\vec{Z}_{deg} - \vec{Z}_{clean})}{\vec{Z}_{clean}} \times 100 \quad (5)$$

$\Delta\vec{Z}$, in the above equation is the measurement deviation vector between the healthy/clean engine sensors data $\vec{Z}_{clean}$, and degraded engine's data i.e., $\vec{Z}_{deg}$.Furthermore, noise was added to the measurement deltas to account for measurement uncertainties that happen in the experimental data. The standard deviation for Gaussian distribution was assumed to be 1% for temperature signals, while 0.5% for pressure signals. The equation involved in the noise generation using random function is as follows,

$$x = -1 + 2 \times rand(1, N) \quad (6)$$

N expresses the number of sample points including clean and faulty engine data. A total of 800 sample points were generated, i.e., 400 samples for each NG and hydrogen fuel scenario.

2.3.2. ANN based classification

After accomplishing preprocessing of the data, data were fed to the ANN based classification learner in form of two separate data sets i.e., NG and hydrogen, using MATLAB tool. Using scenarios involved in the labeled data, a classification algorithm "learns" about classifying fresh observations through a supervised machine learning approach. Although, there are plenty of other algorithms for classification learning, the ANN was chosen in the present study. The ANN architecture is show in Fig. 5. The reason for choosing ANN lies in the inherent ability of this algorithm to (i) capture nonlinear behavior of engine performance efficiently (Fentaye et al. 2019), (ii) extract information in fast and simplistic way (Tahan et al. 2017), (iii) handle multiple and larger component faults in presence of sensors faults (Ogaji and Singh 2003), (iv) deal with measurement uncertainties (Marinai, Probert, and Singh 2004), and (v) perform diagnosis with scarcity in measurements (Singh 2003). In classification learner, a validation method needs to be chosen to assess the prediction accuracy of the fitted model. The validation not only provides performance estimations of the model on completely new dataset (as compared to the training dataset), but also helps in protecting against overfitting. The validation scheme chosen in the present study, however, is k-folds cross validation. This scheme works by dissecting the training datasets into k disjoint sets or partitions and then randomly shuffles them. For each round of training-validation, a certain partition is used for validation while the rest of the data is used for testing. Therefore, each partition is used once for validation while $k-1$, times for training. The k was assumed 5 in the present case based on the data

samples. Cross validation helps in avoiding the overfitting of the training data so that the prediction accuracy might not be compromised.

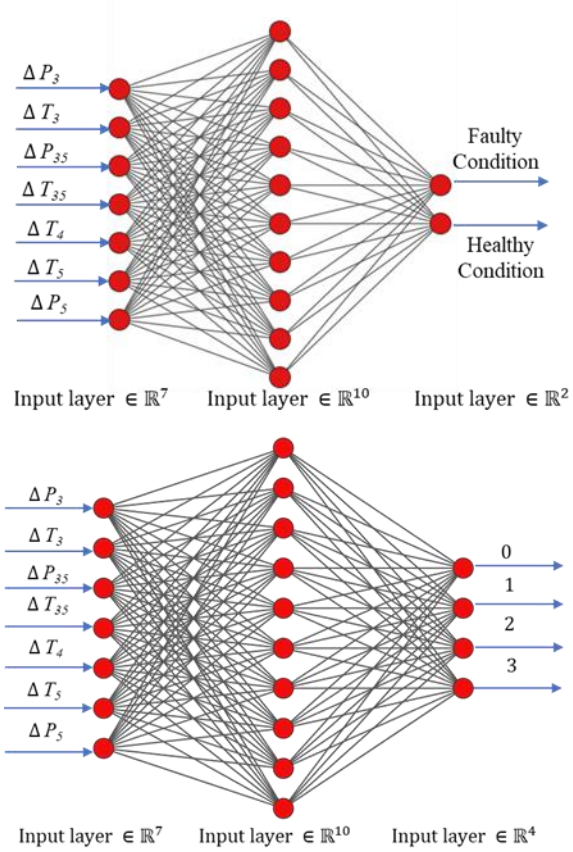


Figure 5: ANN architecture for both fault detection and isolation

The classification algorithms finally provide a confusion matrix that determines the number of faults accurately predicted or wrongly predicted. Confusion matrix provides information about the performance of the selected classifier in each class i.e., *True Class* or *Predicted Class*. The rows in the matrix show *True Class*, while the columns represent *Predicted Class.* The diagonal cells depict the matching of both True and Predicted classes. The blue color in the diagonal cells illustrates that the classifier has classified the observations correctly. The confusion matrix plot is also accompanied with two more separated columns on the far-right hand side that show the performance of the classifier per class in terms of *True Positive Rate (TPR)* and *False Negative Rates (FNR)*. *TPR* is basically the proportion of correctly classified observations per true class while *FNR* shows the proportion of the incorrectly classified observation per true class.

Another way of determining the classifier performance is by observing the results per *Predicted Class* (instead of *True Class*) in terms of Positive Predictive Values (PPV) and False Discovery Rates (FDR). The PPV represents the proportion of correctly classified observations per predicted class. The FDR measures how many observations are classified wrongly for each predicted class. The confusion matrix now has summary rows far below the table when this choice was made. PPV for properly predicted points in each class are displayed in blue, and FDR for erroneously predicted points in each class are displayed in orange.

2.3.3 ANN based fault identification

The final step involved in an MGT diagnosis process is fault identification. The present study utilizes a multi-layer perceptron (MLP) for the intended component fault identification. MLP is a kind of feed forward neural network that works on the concept of supervised learning comprising of input layer, output layer, and one or more hidden layers. In the training phase of the ANN, the network manages to learn the correlations between the input and output data using back propagation algorithm. The current study utilizes a single layer MLP with 10 nodes as shown in Fig. 6. In general, the fault identification is carried out by tracing the health parameters i.e., (Flow capacity: Γ, Efficiency: η) back from the deviated fault signatures. In Fig. 6, on the left-hand side of the ANN structure inputs are provided that have been derived from the equation 5 while the outputs illustrated on right hand side of the structure have been derived from equation 2 and 3. The terms with Δ in the figure represent measurement deviations while Γ and η represents the flow capacity and efficiency of compressor and turbines. The network was trained on the three fault scenarios (CF, TC, CF+TC) to identify some suitable relationships from the fed samples thereby fine tuning the weights and biases. The performance of the training or prediction accuracy is determined by mean square error (MSE) by combining the results from both training and validation data sets. The training progress data and model summary of ANN algorithm have been listed in Tab. 3.

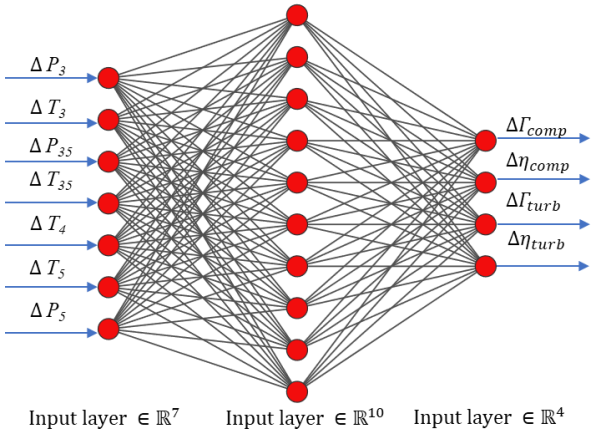


Figure 6: Single layer MLP architecture

Table 3: ANN training progress data and model summary

Criteria	Indicators
Total hidden layers	1
Neurons in hidden layers	10
Feeding approach	Backpropagation
Target limit of epochs	1000
Performance accuracy target	0
Performance gradient target	1.00e-07
Activation function	Sigmoid
Training algorithm	Levenberg-Marquardt
Performance indicator	Mean square error

3. Results and Discussion

3.1. Fault detection and isolation

For fault detection and isolation, ANN based classifier was employed. Two kinds of data sets were trained, and three fault scenarios were accounted for. For each data set, 70% of the data were utilized for training while the remaining 30% were employed for testing and validation (15% for each) of the algorithm. The performance of the classification algorithm is normally assessed by detection decision matrix and classification confusion matrix consisting of the main decision metrics parameters i.e., True Positive (TP), False Negative (FN), False Positive (FP), and True Negative (TN) as illustrated in Fig. 7. The main diagonal depicts correctly predicted faults while off diagonal show wrongly predicted elements. The detection rates of these decision parameters can be estimated through normalization that is done via dividing each matrix' element by sum of its row's elements as follows (Simon 2010),

$$TPR = \frac{TP}{TP + FN} \times 100\% \quad (7)$$

$$FPR = \frac{FP}{FP + TN} \times 100\% \quad (8)$$

$$FNR = \frac{FN}{TP + FN} \times 100\% \quad (9)$$

$$TNR = \frac{TN}{FP + TN} \times 100\% \quad (10)$$

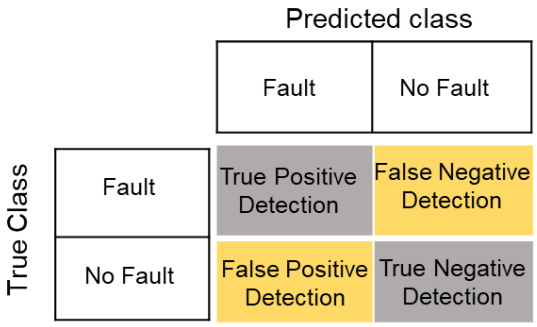


Figure 7: Fault detection decision matrix

The selected classification algorithm enables the classification of multiple faults, as shown in Fig. 8. The figure represents the confusion matrix of NG fueled scenario. The labels mentioned on x- and the y-axis represent different fault and no-fault conditions for predicted and true classes respectively, as listed in Tab. 4.

Table 4: Labels of different physical conditions in the classification algorithm

Label	Designated physical condition
0	No fault
1	CF: Compressor fouling
2	TC: Turbine corrosion
3	CF+TC: Simultaneous

It became evident from Fig. 8 that, at no fault condition, the classifier predicted 99.1% correctly as clean, while 0.9 % wrongly predicted as corrosion. Similarly, in row two, 98.2% data points were truly classified as faulty with compressor fouling, while 1.8% was wrongly predicted as non-faulty. The third row depicts that 98.2% samples were correctly predicted as corroded, while 0.9% wrongly predicted as non-faulty and other 0.9% appeared to be wrongly predicting as simultaneous compressor fouling and turbine corrosion. The fourth row, however, shows that 95.5% of the data points were predicted as representing simultaneous fault (CF+TC), while 1.8% wrongly predicted as fouled, and 2.7% wrongly predicted corrosion. The overall positive prediction value (PPV) as shown in Fig. 8 were found as follows, clean condition: 97.3%, fouled: 98.2%, corroded: 96.4%, and simultaneous fouled and corroded: 99.1%. However, the rest of the data samples showed a false detection rate (FDR).

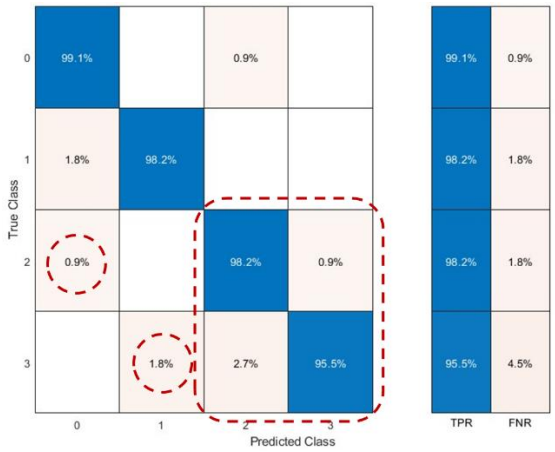


Figure 8:Classification confusion matrix per true class for NG scenario

Figure 9: Classification confusion matrix per predicted class for NG scenario

The fault detection and isolation of hydrogen fuel scenario was conducted similarly as NG scenario. The classification confusion matrix of the hydrogen fuel scenario has been illustrated in Fig.10. At clean condition 0.9% wrong prediction of fouling has been indicated, while 99.1% of the data samples were correctly classified as no fault samples. At fouling condition (depicted in row two), 1.8% samples showed a wrong classification as non-faulty, whereas 98.2% were correctly classified as fouled samples. In the corrosion fault, 3.6% data samples were wrongly classified as non-faulty, while 96.4% showed a correct prediction of corrosion fault. Likewise, at simultaneous CF+TC fault, 93.6% data points indicated the simultaneous fault as correctly, while there was wrong prediction of 1.8% as non-faulty, 1.8% as fouled, and 2.7% as corroded components faults.

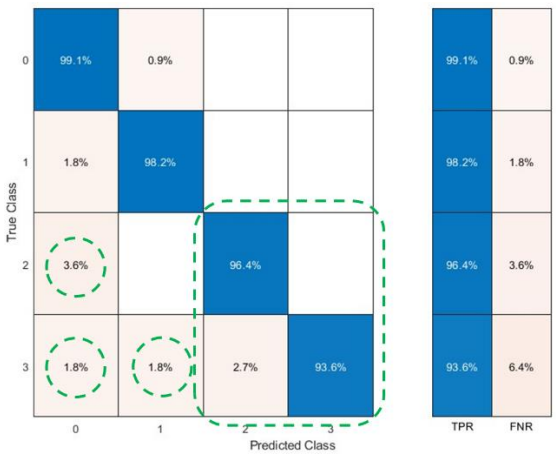


Figure 10: Classification confusion matrix per true class for hydrogen scenario

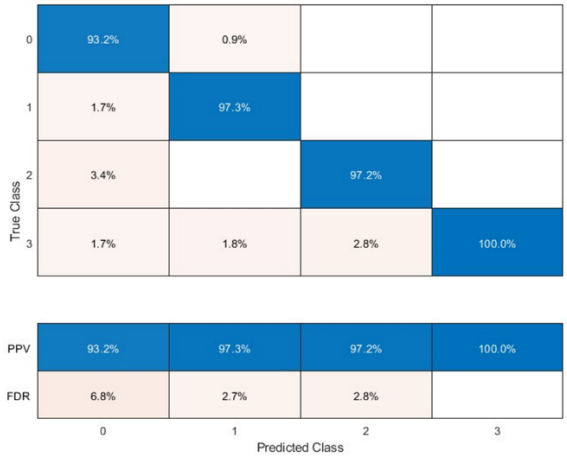


Figure 11: Classification confusion matrix per predicted class for hydrogen scenario

A comparison of Fig. 8 and 10 for corrosion fault shows that the percentage of wrongly classified faults in hydrogen fuel scenario (3.6%) is more than that of natural gas scenario (1.8%). Similarly, for a simultaneous fault, the utilization of hydrogen fuel is stimulating more wrongly classified fault scenarios i.e., one extra wrongly classified prediction of non-faulty case of 1.8%. The three blue and green dotted circles have been drawn on confusion matrix of both NG and hydrogen scenarios respectively. The green circles are showing extra anomalies in the data as compared to NG one. Among the two anomalies i.e., at cell (2,0), and (3,0) are representing deviation in two of the True classes i.e., corrosion and simultaneous CF+ TC. In case of TC, the True class is showing 3.6% wrongly identified faults as non-faulty that is almost double than the NG scenario. Similarly, the simultaneous CF+TC case, is also indicating 1.8% data as wrongly identified as non-faults in presence of the simultaneous CF+T fault. In contrast, in case of NG this wrong non faulty prediction was not observed. However, it is important to mention that the correctly predicted fault rates percentages of a

hydrogen fuel scenario are marginally lower than those of a NG scenario. Additionally, the comparison of Fig. 9 and 11 indicates that PPV of the hydrogen fuel scenario is 100%, while for NG scenario it was 99.1%. It implies that the hydrogen fuel scenario has more propensity of providing positive prediction of faults.

During fault detection, the presence of hydrogen-based corrosion fault led to increased level of incorrectly classified 'no fault' as hydrogen induced corrosion fault. Moreover, the simultaneous compressor fouling, and turbine corrosion faults prompted an extra non faulty prediction in contrast with the NG case. It means presence of the hydrogen induced corrosion fault might influence the fault detection process by giving wrong alarm of fault while there is no actual fault.

3.2 Fault identification

The fault identification of the MGT was carried out by using feed forward neural network. The computational framework of the MLP based ANN is through regression analysis. The regression plots of training, testing and validation have been illustrated in Fig. 12. The figure indicates that ANN was able to identify the physical faults parameters with quite good accuracy since the regression values are almost closer to 1. The performance of the ANN prediction is normally evaluated by level of error minimization with respect to the number of epochs (cycles). An epoch is basically the training process of ANN with all the available data at once for one cycle. It is always desirable to keep accuracy as high as possible during the training.

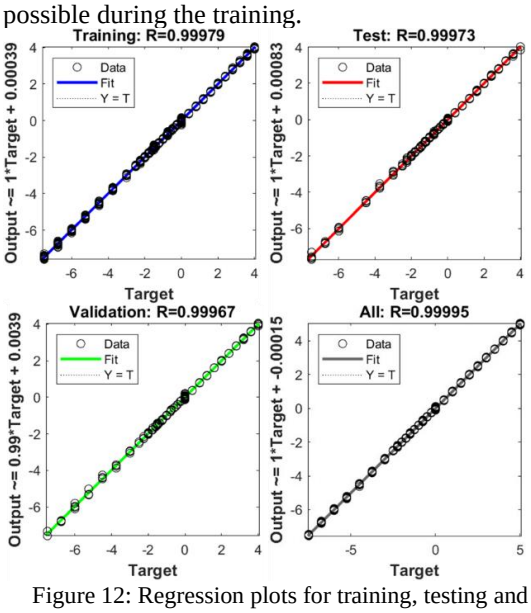


Figure 12: Regression plots for training, testing and validation of a NG fuel ANN model

Normally, a learning curve graph helps in visualizing the convergence of the training, testing and validation; and hence provides information about the accuracy in given epochs. The learning curve graph keeps on getting better until the model coverages with a minimized error, as shown in Fig. 13.

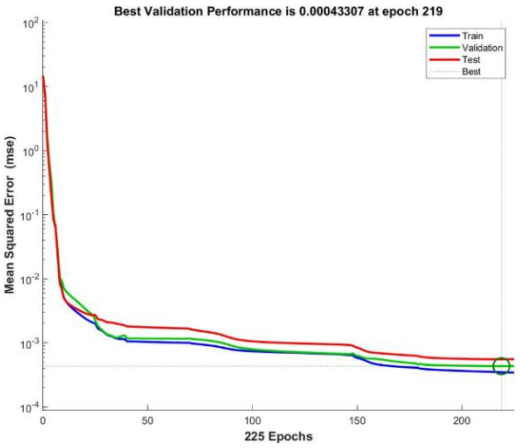


Figure 13: Performance of the ANN training for the NG fuel scenario

The regression of both NG and hydrogen fuel scenarios was found to be nearly identical with similar accuracy as can be observed from Fig. 14. The learning curve-based performance of hydrogen fuel scenario is shown in Fig. 15. The error minimization of the hydrogen scenario took 404 epochs that was significantly greater than that of the NG scenario i.e., 202 epochs.

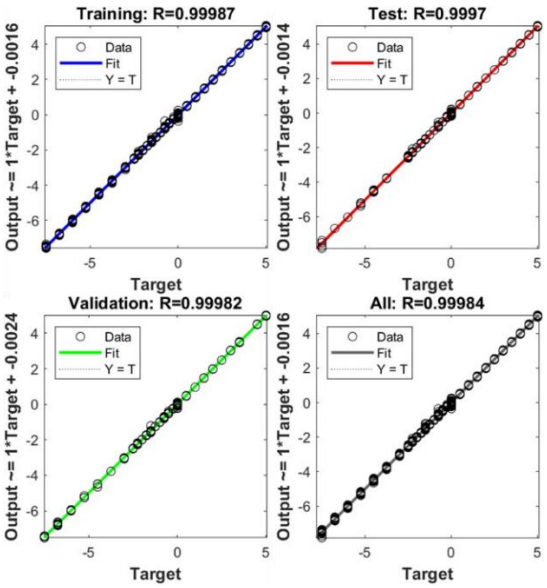


Figure 14: Regression plots for training, testing, and validation of a hydrogen fuel ANN model

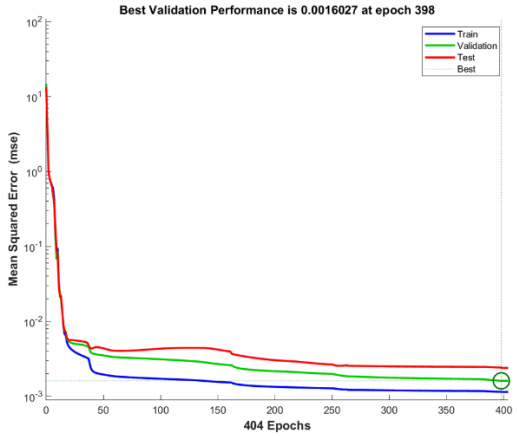


Figure 15: Performance of ANN training for the hydrogen fuel scenario

The overall accuracy of the ANN training, testing and validation was determined by root mean square error (RMSE) using Equation 11. The n, involved in the equation expresses the total sample size, while p_f, and t_f represent the predicted and target fault values. The final results of both MSE and RMSE between the predicted and target values for both NG and hydrogen scenario are listed in Tab. 5. It became evident from the RMSE of the training, testing and validation phases of hydrogen were slightly higher than that of NG scenario.

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(p_f - t_f\right)^2} \qquad (11)$$

Table 5: Analysis of the fault identification results in terms of RMSE

	MSE NG	MSE Hydrogen	RMSE NG	RMSE Hydrogen
Training	5.92e-4	0.0018	0.0243	0.0424
Testing	6.41e-4	0.0022	0.0253	0.04694
Validation	7.40e-4	0.0032	0.0271	0.0565

4. Conclusion

The study was aimed at developing a fault diagnosis model for a hydrogen fueled MGT in comparison with a NG fueled case. The study involved development of a physics-based model for data generation as an initial step. The data further went through preprocessing phase prior to the fault detection, isolation, and identification. Fault detection and isolation were carried out using an ANN based classification learner, while fault identification was performed using an MLP feed forward ANN. The detection and isolation module showed greater percentages of wrongly classified faults due to involvement of steam induced corrosion in hydrogen fueled scenario as compared to a NG fired MGT. The hydrogen scenario showed more propensity of positive prediction values too. The performance of fault identification was however, found to be similar for both NG and hydrogen-based scenarios. Further work is needed with increased level of fault severity rising due to steam induced corrosion for better fault identification in hydrogen fueled scenario. The study was part of an initial attempt towards fault diagnosis of hydrogen fueled micro gas turbines. However, further advancements might help the design and maintenance engineers in assuring optimum reliability and availability of the MGT.

Acknowledgments

The authors would like to thank the University of Stavanger, and the Equinor's Akedemia program for providing the invaluable support in conducting this research. Mohammad Mansouri would like to acknowledge the Robinson project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957752. It should be noted that this paper reflects only the authors' views, and the Research Executive Agency, Equinor, and the European Commission are not liable for any use that may be made of the information contained therein.

References

Bauwens, PL. 2015. 'Gas path analysis for the MTT micro turbine'.

Chen, Yu-Zhi, Xu-Dong Zhao, Heng-Chao Xiang, and Elias Tsoutsanis. 2021. 'A sequential model-based approach for gas turbine performance diagnostics', *Energy*, 220: 119657.

EDGAR/JRC. 2022. "Global carbon dioxide emissions from 1970 to 2021, by sector (in billion metric tons of carbon dioxide)." In.: Statista.

Fentaye, Amare D, Aklilu T Baheta, Syed I Gilani, and Konstantinos G Kyprianidis. 2019. 'A review on gas turbine gas-path diagnostics: State-of-the-art methods, challenges and opportunities', *Aerospace*, 6: 83.

Gazzani, Matteo, Paolo Chiesa, Emanuele Martelli, Stefano Sigali, and Iarno Brunetti. 2014. 'Using hydrogen as gas turbine fuel: premixed versus diffusive flame combustors', *Journal of Engineering for Gas Turbines and Power*, 136.

Gomes, EEB, D McCaffrey, MJM Garces, AL Polizakis, and P Pilidis. 2006. "Comparative analysis of microturbines performance deterioration and diagnostics." In *Turbo Expo: Power for Land, Sea, and Air*, 269-76.

Kim, Sangjo, Ju Hyun Im, Myungho Kim, Junghoe Kim, and You Il Kim. 2023. 'Diagnostics using a physics-based engine model in aero gas turbine engine verification tests', *Aerospace Science and Technology*: 108102.

Kurzke, J. 2012. "GasTurb 12: A program to calculate design and off-design performance of gas turbines. User's manual." In.: GasTurb, Aachen, Germany.

Marinai, Luca, Douglas Probert, and Riti Singh. 2004. 'Prospects for aero gas-turbine diagnostics: a review', *Applied energy*, 79: 109-26.

Noble, David, David Wu, Benjamin Emerson, Scott Sheppard, Tim Lieuwen, and Leonard Angello. 2021. 'Assessment of current capabilities and near-term

availability of hydrogen-fired gas turbines considering a low-carbon future', *Journal of Engineering for Gas Turbines and Power*, 143.

Oluyede, Emmanuel O, and Jeffrey N Phillips. 2007. "Fundamental impact of firing syngas in gas turbines." In *Turbo expo: power for land, sea, and air*, 175-82.

Tahan, Mohammadreza, Elias Tsoutsanis, Masdi Muhammad, and ZA Abdul Karim. 2017. 'Performance-based health monitoring, diagnostics and prognostics for condition-based maintenance of gas turbines: A review', *Applied energy*, 198: 122-44.

Talebi, SS, A Madadi, AM Tousi, and M Kiaee. 2022. 'Micro Gas Turbine fault detection and isolation with a combination of Artificial Neural Network and off-design performance analysis', *Engineering Applications of Artificial Intelligence*, 113: 104900.

Talebi, SS, and AM Tousi. 2017. 'The effects of compressor blade roughness on the steady state performance of micro-turbines', *Applied Thermal Engineering*, 115: 517-27.

TURBEC, SPA. 2017. "Technical Description Microturbine Turbec T100. people. unica. it/danielecocco/files/2012/07/Microturbina_ T100_Detailed_Specifications1. pdf." In.: Stand.

TURBINE, ZERO-CARBON GAS. 'HYDROGEN GAS TURBINES'.

Urban, Louis A. 1975. 'Parameter selection for multiple fault diagnostics of gas turbine engines'.

Urban, Louis A, and Allan J Volponi. 1992. 'Mathematical methods of relative engine performance diagnostics', *SAE Transactions*: 2025-50.

Ying, Yulong, and Jingchao Li. 2023. 'An improved performance diagnostic method for industrial gas turbines with consideration of intake and exhaust system', *Applied Thermal Engineering*, 222: 119907.

Yoon, Jae Eun, Jong Jun Lee, Tong Seop Kim, and Jeong L Sohn. 2008. 'Analysis of performance deterioration of a micro gas turbine and the use of neural network for predicting deteriorated component characteristics', *Journal of Mechanical Science and Technology*, 22: 2516.

Zwebek, A, and P Pilidis. 2003. 'Degradation Effects on Combined Cycle Power Plant Performance—Part II: Steam Turbine Cycle Component Degradation Effects', *J. Eng. Gas Turbines Power*, 125: 658-63.

Zwebek, AI, and P Pilidis. 2004. 'Degradation effects on combined cycle power plant performance—Part III: Gas and steam turbine component degradation effects', *J. Eng. Gas Turbines Power*, 126: 306-15.

Hydrodynamic study of a CO_2 desorption column using computational fluid dynamics.

Sumudu Karunarathne*, Kristoffer Eikeseth, Lars Erik Øi

University of South-Eastern Norway, N-3901 Porsgrunn, Norway

sumudu.karunarathne@usn.no

Abstract

Desorption of CO_2 from the rich amine solvent is one of the main operations in the amine-based CO_2 capture process. Proper vapour and liquid flow through the packing materials would enhance the heat transfer that is needed for stripping CO_2 from solvent. This is achieved by increasing the surface area of the flowing solvent by using the packing material. In this study, the created CFD (Computational Fluid dynamics) model in OpenFOAM^{TM} was able to simulate the factors influencing TCM (Technology Centre Mongstad) desorption performance, including liquid distribution, wettability and film thickness within the packing material. Three scenarios were considered including a base case for a better understanding of the hydrodynamics in the desorption column. Two of these are to compare the influence of mass flow rates, while one is used to investigating potential improvement. Simulation revealed that introducing a deflector plate and CO_2 bypass tube has a positive hydrodynamic effect in the desorption column.

1. Introduction

Carbon dioxide (CO_2) capture using amine solvents is a matured technology and has been used in the natural gas industry for decades. Various research has been performed to investigate the feasibility of employing the technology in post-combustion CO_2 capture. Technology Centre Mongstad (TCM) in Norway is a test facility that enables examining new solvents and to perform many other CO_2 capture related activities.

The process of post-combustion CO_2 capture undergoes cyclic absorption and desorption of CO_2 as illustrated in Fig. 1. The 30% wt monoethanolamine (MEA) is a benchmark solvent that has been tested many times to explore its capabilities for capturing CO_2 (Martinez, *et al*. 2017). The efficiency of absorption and desorption depends on the mass and heat transfer in the absorption and desorption columns. Thus, hydrodynamic of gas/vapour and liquid through packing material plays a vital role has a greater influence on mass and heat transfer between gas and liquid phases. Ideally, the solvent flows through the packing material and spreads out as an even, thin film across the entire packing material surface. However, several factors may negatively influence the flow, including:

- Velocities and distribution of liquid as it enters the packing bed.
- Angled channels in the packing material may force flow against the desorption tower walls, creating thick and fast flowing channels of solvent.
- CO_2 rising through the column in a turbulent manner may disturb the wetted surfaces.
- Liquid holdup may occur, creating localized flooding.

The aim of this work was to create a CFD (Computational Fluid Dynamic) model able to simulate factors influencing TCM desorption column performance, including liquid distribution, wettability and film thickness within the packing material. A scaled down version of the actual geometry of TCM desorption column was considered in the simulations.

2. Literature

Computational fluid dynamic studies on gas and liquid flow through packing have been reported in literature. This section cites some of the work done in this field regardless of whether it is absorption or desorption. Niegodajew and Asendrych, 2016 discussed a CFD simulation of small laboratory test rig which has a random packed bed. A 2-fluid Eulerian model has been employed to determine the flow behavior and validated using the reference data from rig. Pham *et al.*, 2015 described an approach taken to simulate an absorber considering the complex structured packing geometry to a homogeneous porous material. The study revealed that the porous media CFD model could reflect hydrodynamics and gas–liquid interactions of structured-packings. Gbadago *et al.*, 2020 performed a CFD simulation of a packed bed industrial absorber with interbed liquid distributors.

A porous media was used to represent the packing material in CFD simulations to avoid the high computational cost associated with simulating a real structured packing.

Yang *et al.*, 2018 presented a CFD based column study in which Mellapak 250 Y was selected as the packing material in geometry due to its thorough characterization. A similar analysis was performed by Isoz, 2017 by considering the real geometry of Mellapak 250.X and Mellapak 250.Y to study the gas flow through the structured packing. Raynal *et al.*, 2004 showed the possible ways CFD can be used for hydrodynamics calculations, liquid holdup and pressure drop within structured packings. Haroun *et al.*, 2012 extended the work by including a computational analysis of mass transfer in structured packings.

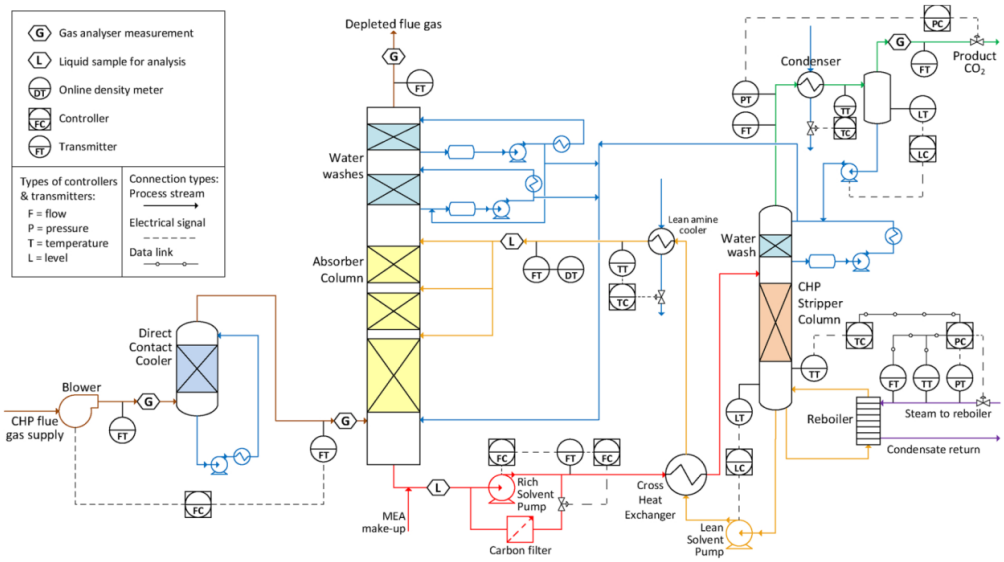


Figure 1: Process flow diagram of the TCM CO_2 capture facility. (Bui *et al.*, 2020)

3. Methodology

3.1. Geometry and Mesh

The geometry and mesh were created by using software tools Blender and SnappyHexMesh respectively. The simulation has been scaled down to make it viable to simulate with the available computing power. Symmetry is assumed to acquire a good representation of flow through the packing material and wall effects. The wall circumference to surface area ratio is not equal to the desorber design at TCM. This introduces some uncertainties in the study of hydrodynamic wall interface actions. The desorber at TCM has a packing bed height of 8 meters, has 28 packing bed layers and is 1.25 meters wide with 108 liquid distribution points. The performed simulation has been scaled down to a height of 0.92 meters with 3 packing bed layers and a width of 0.12 meters with 1 liquid distribution point. Fig. 2 illustrates a sectioned view of created geometry for simulations.

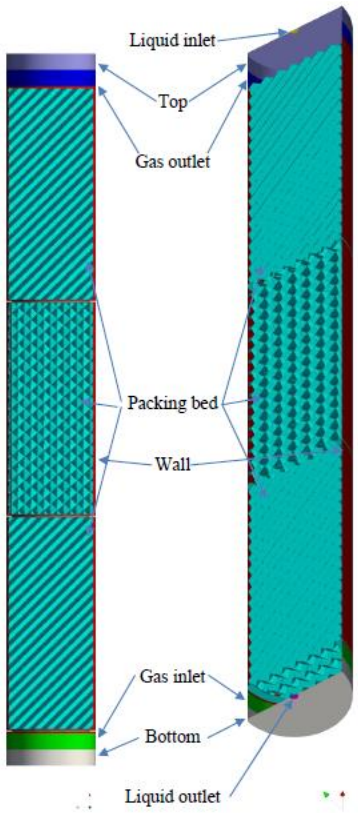


Figure 2: Sectioned view showing color-coded geometry, left = axis normal. Right = isometric view.

The mesh was created in two stages. First the geometry was created in Blender using a script created by Isoz, 2017 and modified to facilitate two-phase flow and geometries of the TCM desorption tower.
The SnappyHexMesh was used to create 3 sizes/types of cells as indicated below and matched them to the geometry as shown in Fig. 3,4 and 5.

- Layer 0 cells: cells further than 1.5 mm away from any surface, start as 1.5 mm hexahedral cells, but are formed and shaped to fit geometry as SnappyHexMesh runs.
- Layer 1 cells: cells closer than 1.5 mm to the surface, starting as 1.5 mm hexahedral cells, but are split into eight 0.75 mm hexahedral cells and formed and shaped to fit geometry as SnappyHexMesh runs.
- Surface layer cells: Added to surfaces of cylinder walls and packing material to achieve better results in simulations for gradients in velocity and film thickness. Two surface layers are added to the packing material, and one is added to the cylinder walls. Layer 1 and layer 0 cells are pushed back to accommodate surface layers as a surface layer addition, which is the last step of mesh generation in SnappyHexMesh.

The final mesh has 11.68 million cells in the layer 0 and 1 and 4.98 million cells in the surface layer with a total of 16.66 million cells. A mesh independence analysis has not been performed, but previous analysis on similar geometries has found 4 million cells per layer of packing material to be sufficient Isoz, 2017. These simulations have been performed with roughly 5.55 million cells per packing element.

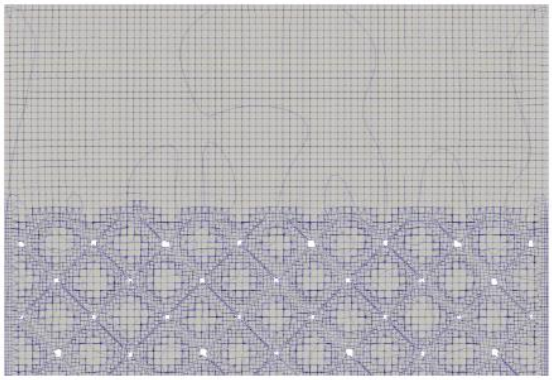


Figure 3: Sectioned view showing mesh at the top.

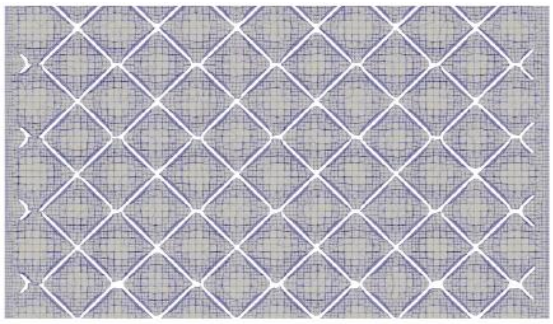


Figure 4: Sectioned view showing mesh in the middle.

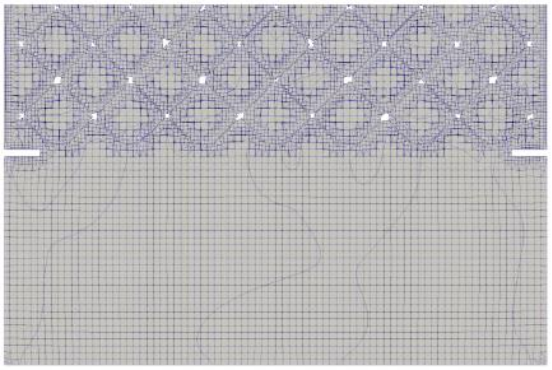


Figure 5: Sectioned view showing mesh at the bottom.

3.2. Mathematical models

The simulations were performed using the solver called InterFoam in OpenFOAM. It is a widely used solver for multiphase simulations of two incompressible, isothermal immiscible fluids in which VOF (volume of fluid) phase-fraction based interface capturing approach is adopted for computations (Heyns and Oxtoby, 2014).

Continuity Equation:

$$\frac{\partial u_j}{\partial x_j} = 0 \qquad 01$$

Where, u is velocity.

Momentum Equation:

$$\frac{\partial(\rho u_i)}{\partial t} + \frac{\partial(\rho u_j u_i)}{\partial x_j} = -\frac{\partial P}{\partial x_i} + \frac{\partial}{\partial x_j}\left(\tau_{ij} + \tau_{t_{ij}}\right) + \rho g_i + f_{\sigma i} \qquad 02$$

Where, g is gravitational acceleration, ρ is density, P is pressure, τ_{ij} is viscous stress, $\tau_{t_{ij}}$ is turbulent stress and $f_{\sigma i}$ is surface tension.

$$\rho = \alpha\rho_1 + (1-\alpha)\rho_2 \qquad 03$$

Here, α is 1 inside fluid 1 with the density ρ_1 and 0 inside fluid 2 with the density ρ_2. At the interphase between the two fluids α varies between 0 and 1.

The surface tension $f_{\sigma i}$ is modelled as continuum surface force (CSF) and calculated as follows.

$$f_{\sigma i} = \sigma\kappa\frac{\partial\alpha}{\partial x_i} \quad 04$$

σ is the surface tension constant and κ the curvature. The curvature can be approximated as follows:

$$\kappa = -\frac{\partial n_i}{\partial x_i} = -\frac{\partial}{\partial x_i}\left(\frac{\partial\alpha/\partial x_i}{|\partial\alpha/\partial x_i|}\right) \quad 05$$

Equation for the interphase:
An additional equation for α has to be solved in order to know where the interphase between the two fluids is,

$$\frac{\partial\alpha}{\partial t} + \frac{\partial(\alpha u_j)}{\partial x_j} = 0 \quad 06$$

The equation can be seen as the conservation of the mixture components along the path of a fluid parcel.

3.3 Simulations

Three simulation cases have been considered to investigate the desorber performance and archive a good understanding of the hydrodynamics. Tab. 1 lists the parameters considered in each simulation case while Tab. 2 provides thermophysical properties of CO_2 and MEA solvents.

Table 1: Specifications of the simulation cases.

Parameter	Base[a]	ICL07[a]	Defl_Exp
CO_2 mass flow rate (g/s)	7.70	8.07	4.00
MEA mass flow rate (g/s)	125.57	109.47	109.47
Geometry	TCM scale-down	TCM scale-down	TCM scale-down with added MEA Deflector plate

Reference (a): Bui *et al.*, 2020.

Both Base case and ICL07 have scaled-down values from plant operating parameters (Bui *et al.*, 2020). Base case is used as a reference for normal operating parameters, which is compared to ICL07, chosen because it gives a high CO_2 capture rate (Bui *et al.*, 2020).
The Delf_Exp is an experimental case where two potentially efficiency increasing factors were introduced:

i. A deflector plate is introduced into the stream of MEA before it enters the packing bed, the deflector plate may help in distributing MEA before it enters the packing bed, the MEA solvent may also become smaller droplets before entering the packing bed, thus releasing more CO_2 before entering the packing bed.
ii. The mass flow of CO_2 counterflowing MEA solvent through the packing bed is reduced to lower gas flow influence on liquid flow. In practice this might be accomplished by introducing a bypass pipe for CO_2 running parallel with the desorption tower along the top half of the packing bed.

Table 2: Thermophysical properties of CO_2 and MEA solvent.

Property	CO_2	Solvent
Temperature [K]	383	383
Density [kg/m^3]	1.384	1057[a]
Kinematic Viscosity [m^2/s]	1.37×10^{-5}	9.08×10^{-7}
Dynamic Viscosity [Pa·s]	1.83×10^{-5}	9.6×10^{-4} [b]
CO_2 loading (mol_{CO2}/mol_{MEA})		0.51[c]
MEA concentration (wt%)		27.5[c]

Reference (a): Han *et al.*, 2012; (b): Arachchige *et al.*, 2019; (c): Bui *et al.*, 2020.

The simulations were set to end at 7 seconds and results were taken at the 7th second as the final outcome in the simulation. All simulations were performed on virtualized Amazon AWS servers with 64 ARM cores of an AWS Gravitron3 CPU and 128 GB DDR5 RAM, giving a runtime of 72 hours per simulation.

3.4 Boundary conditions

Boundary conditions are necessary to solve governing equations. The solver used, InterFoam requires the following boundary conditions to be defined: Velocity (u), pressure (p), temperature (T) and turbulent volume fraction (α_{MEA}). The turbulence model, k-ε, requires the following boundary conditions to be defined: Turbulent kinetic energy (k), turbulent energy dissipation rate (ε) and turbulence viscosity (μ_t). Boundary conditions used are listed in Tab. 3 and 4.

4. Results

The pressure drop across the packed bed was examined in the different simulation cases. As shown in Fig. 6, the "base" case has the highest pressure drop of 605 Pa or 705 Pa/m. The simulation case "ICL07" has a pressure drop of 535 Pa or 624 Pa/m and "Defl_Exp" the lowest pressure drop of 366 Pa or 427 Pa/m.

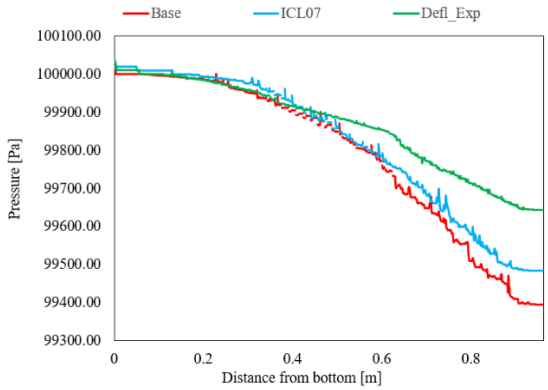


Figure 6: Comparison of pressure drop between cases.

It is remarked that even though "Delf_Exp" has half the gas flow of "ICL07", the pressure drop is not halved. Also, "Delf_Exp" has a smoother pressure graph, indicating low liquid hold-ups. This pressure drop is further illustrated in Fig.7 where the pressure is higher at the bottom and lower at the top of the column.

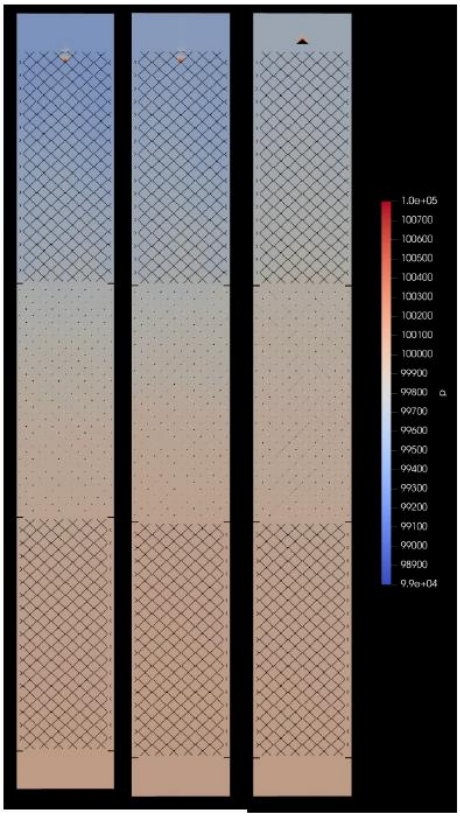


Figure 7: Sectioned view showing pressure, Left = "Base", Middle = "ICL07", Right = "Defl_Exp".

The simulated Base case and ICL07 case indicated severe liquid hold-up with large gatherings of liquid at certain points of the packing bed, while Defl_Exp case has some build-up, but it is not severe compared to other two cases as shown in Fig. 8. A smaller amount of liquid present in the bottom layer indicates the necessity of a longer runtime than 7 seconds to achieve an equilibrium condition in the simulation.

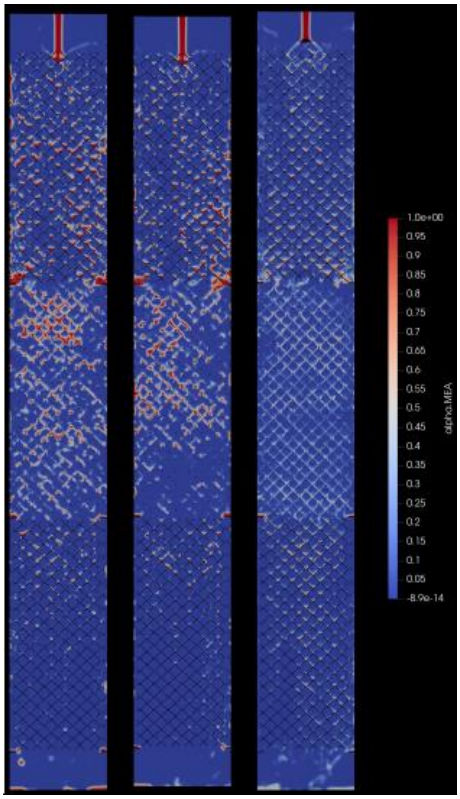


Figure 8: Sectioned view showing solvent distribution, Left = "Base", Middle = "ICL07", Right = "Defl_Exp".

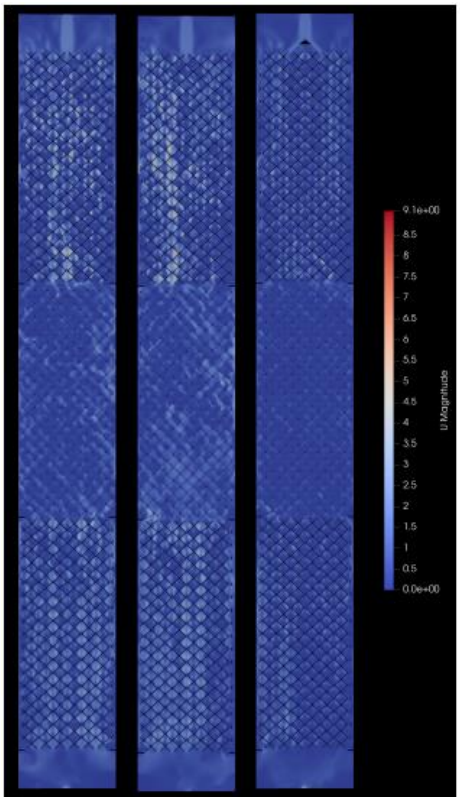
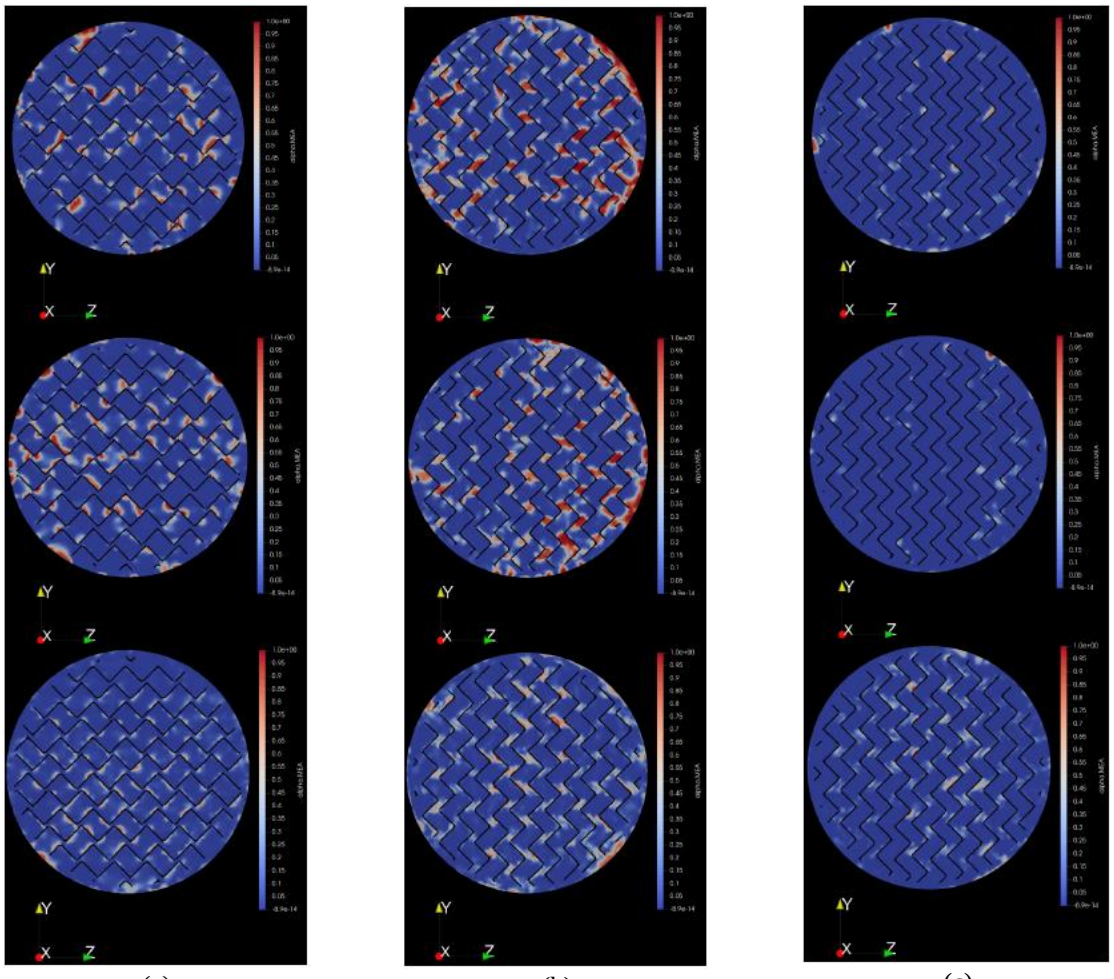


Figure 9: Sectioned view showing velocity, Left = "Base", Middle = "ICL07", Right = "Defl_Exp".

Fig. 9 illustrates the variation of gas velocity along the packed bed. A localized increase of velocity was observed in regions where less amount of liquid flows through the packing.

Table 3: Patch boundary conditions.

Variable	Liquid Inlet	Liquid Outlet	Gas Inlet	Gas outlet
u	flowRateInletVelocity constant	matchedFlowRateOutletVelocity Matched: Liquid Inlet	flowRateInletVelocity constant	matchedFlowRateOutletVelocity Matched: Gas Inlet
p_rgh	fixedFluxPressure	prghTotalPressure uniform 100000	fixedFluxPressure	fixedFluxPressure
T	fixedValue uniform 383	zeroGradient	zeroGradient	zeroGradient
k	fixedValue uniform 0.01	inletOutlet	fixedValue uniform 0.01	inletOutlet
ε	fixedValue uniform 0.02	inletOutlet	fixedValue uniform 0.02	inletOutlet
α_{MEA}	fixedValue uniform 1	zeroGradient	fixedValue uniform 1	zeroGradient
μ_t	calculated	calculated	calculated	calculated



(a) (b) (c)

Figure 10: (a) Sectioned view showing MEA distribution at midpoint in the middle distribution bed layer, Top = “Base”, Middle = “ICL07”, Botom = “Defl_Exp”. (b) Sectioned view showing MEA distribution at midpoint in the top distribution bed layer, Top = “Base”, Middle = “ICL07”, Bottom = “Defl_Exp”. (c) Sectioned view showing MEA distribution at midpoint in the bottom distribution bed layer, Top = “Base”, Middle = “ICL07”, Botom = “Defl_Exp”.

Table 4: Wall boundary conditions

Variable	Top	Cylinder	Packing Element	Bottom
u	noSlip	noSlip	noSlip	noSlip
p	fixedFlux Pressure	fixedFluxPressure	fixedFluxPressure	fixedFluxPressure
T	zeroGradient	zeroGradient	zeroGradient	zeroGradient
k	kqRWall Function	kqRWallFunction	kqRWallFunction	kqRWallFunction
ε	epsilonWallFunction	epsilonWallFunction	epsilonWallFunction	epsilonWallFunction
α_{MEA}	zeroGradient	zeroGradient	zeroGradient	zeroGradient
μ_t	nutkWall Function	nutkWallFunction	nutkWallFunction	nutkWallFunction

Fig. 10 confirms the phenomenon described in Fig.8 that Base case and ICL07 case show severe liquid hold-up with large gatherings of liquid at certain points of the packed bed. Fig. 8 illustrates the liquid hold-up in Defl_Exp case, which shows building up liquid in some regions, but it is not severe compared to the other two cases.

5. Conclusion

This work simulated the top 3 layers out of a total of 28 layers of the TCM desorption column. The performed simulations showed liquid hold-up in the top 3 layers of the TCM desorption column packing bed when it was operated under normal operating parameters. Further it showed that introducing a deflector plate and/or a CO_2 bypass reduces simulated liquid hold-up, but the effect of deflector plate or CO_2 bypass had not been analyzed individually.

As the MEA flows down and releases CO_2 there is a decrease in both MEA and CO_2 mass flow, while also the distribution point(s) has less influence as the liquid flows through the 28 layers in the packing bed. It is therefore likely that the optimal operating conditions for one point in the desorption column yields suboptimal conditions above/below that point.

The simulated results were not verified using experimental data. Accordingly, such investigation is proposed as a future work by using publicly available plant operating data of TCM.

Acknowledgment

The authors gratefully acknowledge the staff of Technology Center Mongstad DA, Gassnova, Equinor, Shell and TotalEnergies for their interest in this work and particularly for access to data from the TCM DA facility.

References

Arachchige, U. *et al.* (2019) 'Dynamic Viscosity of Partially Carbonated Aqueous Monoethanolamine (MEA) from (20 to 150) °C', Applied Chemical Engineering. 2 (2). doi: 10.24294/ace.v1i2.660.

Bui, M. *et al.* (2020) 'Demonstrating flexible operation of the Technology Centre Mongstad (TCM) CO_2 capture plant', International Journal of Greenhouse Gas Control, 93, p. 102879. doi: 10.1016/j.ijggc.2019.102879.

Gbadago, D. Q. *et al.* (2020) 'CFD simulation of a packed bed industrial absorber with interbed liquid distributors', International Journal of Greenhouse Gas Control. 95, p. 102983. doi: 0.1016/j.ijggc.2020.102983.

Han, J. *et al.* (2012) 'Density of Water (1) + Monoethanolamine (2) + CO_2 (3) from (298.15 to 413.15) K and Surface Tension of Water (1) + Monoethanolamine (2) from (303.15 to 333.15) K', J. Chem. Eng. Data. 57(4), pp. 1095–1103. doi: 10.1021/je2010038.

Haroun, Y. *et al.* (2012) 'Mass transfer and liquid hold-up determination in structured packing by CFD', Chemical Engineering Science. 75, pp. 342–348. doi: 10.1016/j.ces.2012.03.011.

Heyns, J. A. and Oxtoby O.F. (2014) 'Modelling surface tension dominated multiphase flows using the VOF approach' 2014. [Online].Available:https://api.semanticscholar.org/CorpusID:16036879.

Isoz, M. 'CFD Study of Gas Flow Through Structured Separation Columns Packings Mellapak 250.X and Mellapak 250.Y', Topical Problems of Fluid Mechanics, 2017.

Martinez, M. *et al.* (2017) 'Solvent selection and design for CO_2 capture – how we might have been missing the point' Sustainable Energy and Fuels. 1(10), pp. 2078-2090. doi.org/10.1039/C7SE00404D.

Niegodajew, P. and Asendrych, D. (2016) 'Amine based CO_2 capture – CFD simulation of absorber performance', Applied Mathematical Modelling. 40(23), pp. 10222–10237. doi: 10.1016/j.apm.2016.07.003.

Pham, D. A. *et al.* (2015) 'Porous media Eulerian computational fluid dynamics (CFD) model of amine absorber with structured-packing for CO_2 removal', Chemical Engineering Science. 132, pp. 259–270. doi: 10.1016/j.ces.2015.04.009.

Raynal, L. *et al.* (2004) 'Liquid Holdup and Pressure Drop Determination in Structured Packing with CFD Simulations', The Canadian Journal of Chemical Engineering. 82, pp. 871-879.

Yang, L. *et al.* (2018) 'CFD Modeling on Hydrodynamic Characteristics of Multiphase Counter-Current Flow in a Structured Packed Bed for Post-Combustion CO2 Capture', Energies, 11, 3103. doi: 10.3390/en11113103.

Information extraction from operator interface images using computer vision and machine learning

Eirik Illing [a,*], Nils-Olav Skeie [b], Ole Magnus Brastein [c]

[a] *Emerson Automation Solutions,* [b c] *University of South-Eastern Norway (USN)*
eirik.illing@emerson.com

Abstract
In the process of system upgrades or migrations, the utilization of existing layouts and object structures for designing new Human Machine Interfaces (HMI) can significantly save time and effort. Operator interface images, commonly referred to as HMI´s, contain valuable information crucial to industrial operations, but access to source code or design files can be limited. Modern frameworks for object detection and text recognition offer a solution by extracting information directly from images. However, these methods require time-consuming data acquisition and manual effort to initiate. This paper proposes a novel approach utilizing traditional Computer Vision (CV) and Machine Learning (ML) techniques to extract objects from images. The extracted objects are used as training data to transfer learn a ResNet model for multi-label image classification. The combination of this model with techniques such as sliding window, pyramid scaling, and non-maximum suppression forms the basis for a semi-automated annotation tool. This tool generates training data for more optimized object detection methods, specifically the YOLO (You Only Look Once) one-stage object detector. The semi-automated annotation tool allows engineers to manually refine the training data and export state-of-the-art training images for YOLO. The YOLO model achieves an impressive mean Average Precision at IoU 50% (mAP^{50}) score of 95.5% when transfer learned on the annotated data. Additionally, an Optical Character Recognition (OCR) engine is utilized to extract text information from preprocessed images, followed by postprocessing to filter tag data. An algorithm is then employed to link objects and tags together. The final solution is implemented in software designed to optimize user interaction, resulting in an analysis document in Excel format, which can be easily exported for end-user access. With the novel use of this software to automate image analysis, the time required to analyze HMI images prior to migration or rebuild can be reduced by an estimate of 90%.

1. Introduction

The rapid advancement of technology has led to an increasing reliance on operator interface images, such as HMI [1] and Supervisory Control and Data Acquisition [2] (SCADA) graphics**,** in various industries. As the field advances the frameworks for these interface technologies evolves, new and improved design concepts are introduced, and migration from old systems to new become a necessity. These traditional operator interface images contain a wealth of valuable information related to production, process flows, and assembly lines. However, accessing the underlying source code or design files of these operator interface images can often be challenging or limited. To address this issue, modern frameworks for image classification and object detection have emerged as potential solutions, enabling the extraction of pertinent information directly from these operator interface images.

This paper explores the field of image classification [3] and object detection [4] for the purpose of extracting information from complex operator interface images. The primary objective is to develop a tool that can effectively analyze and interpret industrial applications depicted in these images. Specifically, the project will be conducted in two iterations, each with distinct goals and outcomes.

1.2. Previous work

Several studies have investigated the recognition and extraction of information from industry related documentation, particularly in the context of Piping and Instrumentation Diagrams (P&IDs). Paliwal et al. [5] proposed a method in 2021 that utilized a Dynamic Graph Convolutional Neural Network (DGCNN) to recognize line-drawn symbols in P&IDs. Their approach involved constructing a graph based on sampled pixels along contour boundaries and incorporating ResNet-34 embeddings to improve classification accuracy. They employed an Arcface loss function to address misclassification issues caused by similar-looking objects [6]. The presented research utilizes a single image for each object, setting it apart from conventional approaches that typically require multiple images for training.

In an earlier paper, Paliwal et al. [7] developed an end-to-end data extraction system for P&IDs using fully convolutional networks. This involved annotating multiple training images with segmented pixels to identify different symbol classes. The authors employed a pipeline approach, separating text extraction and graphic object

detection, and used minimum Euclidean distance to link text and objects. Their system achieved effective information extraction and demonstrated the potential of performing extraction in multiple steps or iterations.

Moon et al. [8], proposed a three-step method for recognizing line objects and flow arrows in image-format P&IDs. Their approach involved removing outer borders and title boxes (considered noise), then detecting continuous lines, line signs, and flow arrows, and adjusting and merging lines accordingly. They employed preprocessing techniques to remove noise, applied thinning and pixel processing for line detection, and utilized a RetinaNet model to train on the line signs and flow arrows.

These studies have made significant contributions to the field of analyzing documentation, specifically in addressing challenges related to object recognition, noise reduction, and information extraction. These studies primarily focused on analyzing documentation in grayscale, predominantly using standardized symbols and texts. In contrast, the current project aims to tackle similar challenges while dealing with operator interface images that exhibit a wide variety of complexities, such as color variations, scales, and limited features, requiring different approaches for object recognition and information extraction.

1.3. Outline of paper

System Description chapter explains how the project was executed, including the system overview and associated advantages and challenges.

Methods chapter covers data collection, training models, and tool development in detail.

Results and Discussion chapter present project outcomes, including model performance, effectiveness of data annotation tools, and discussions regarding the aforementioned topics.

Future Work chapter explores improvement areas, use cases, and opportunities for further development.

Conclusion chapter summarizes key findings and highlights the project's significance and potential impact.

This paper is based on a Master's Thesis project [9] conducted at USN and supported by Emerson Automation Solutions.

2. System description

2.1. Project execution

In the first iteration, the project will focus on training a ResNet50 [10] image classification model. The model will be trained on custom data extracted from existing operator interface images using CV [11-12] techniques, manually sorted into class folders for data labeling [13]. This phase aims to explore different model configurations and assess their performance in accurately classifying industrial objects. An annotation tool is developed to aid in the labeling of training data. This tool will streamline the annotation process and lay the foundation for subsequent iterations.

The second iteration will leverage the annotated data generated from the previous phase to train a one-stage [14] YOLOv8 [15-16] model for object detection. The goal is to accurately identify and locate objects within the operator interface images. Furthermore, text extraction techniques such as Pytesseract OCR [17] will be employed to retrieve textual information from these images, which will then be linked to the respective objects detected using Minimum Euclidean Distance calculations. The text extraction will require extensive image preprocessing to ensure high acquisition accuracy, and a custom text-format library is embedded in the software to filter relevant information. To ensure usability and accessibility, the solutions developed in this project will be combined into one software solution tool that is accessible through a web interface. All development is done using Python and the Flask web framework.

Users will be able to upload operator interface images to the tool, which will then process the images and generate an analysis document as output. This document will provide a representation of the information extracted from the images, facilitating informed decision-making and analysis for industrial applications. To illustrate the benefits and advancements offered, the flowchart provided in Figure 1 compares the proposed analysis process with conventional approaches involving source-code tools or manual analysis.

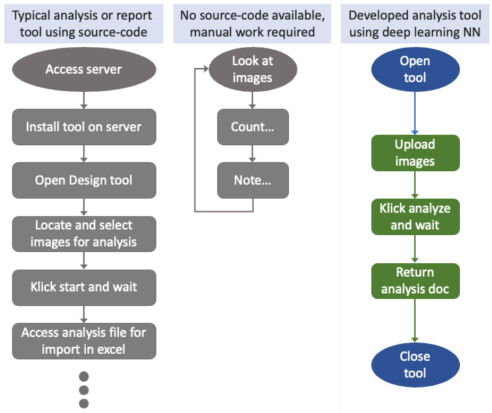


Figure 1: Comparing conventional approaches using source-code tools or manual analysis with the object detection tool suggested in this research.

2.2. Advantages

Performing object detection on computer drawn images, such as drawings and documentations, is in

some ways easier than real-world images. The traditional challenges associated with factors [18] like lighting conditions, object angles, line of sight, dirt, and other real-life variables are non-existent in these 2D images. However, these types of images also present unique challenges that need to be addressed.

2.3. Challenges

Operator interface images contain a large number of objects, lines, and text that represent various types of information. This abundance of visual elements introduces challenges related to noisiness and limited features. Due to the similarities between objects, there is a higher probability of misclassification. In general, there is a lack of true standardization in both object and image design, as well as tag structure within these images.

3. Methods

3.1. Data collection

If source-code or design files is unavailable, extracting training data objects directly from raw image files may be necessary. To streamline this process, a Python script is developed for object extraction, eliminating the need for manual snipping tool usage. OpenCV [19] offers various methods simplifying the extraction of training, validation, and test set objects from raw operator interface images.

The script will convert input images to grayscale, apply thresholding and dilation, and identify contours using the "find contours" method in OpenCV**.** These contours are then enclosed in bounding boxes using a different method from the OpenCV library. Each bounding box represented an object and is snipped from the full-scale image into a separate folder. While some of these objects are suitable for training, validation, and test sets, others are incomplete or contain noise. Although this method is not perfect, it provides a foundation for the subsequent manual sorting of objects into class folders for labeling purposes.

Following the extraction of individual objects from the full-scale images using the Python script, a manual process is undertaken to organize and label the extracted objects. This involves moving the extracted objects into separate folders, with each folder representing a different object class. By separating the objects into distinct folders, it becomes simpler to manage and track the defined classes. Additionally, labeling is achieved by associating each object within a class folder with its corresponding label. While going through the extracted data it is important to check that the objects put into class folders represents the data that the model needs to learn. An example of a clear represented object with the label “valve_p” for valve pneumatic is shown in Figure 2.

Figure 2: Clean pneumatic valve object.

The objects are separated into single-label and multi-label folders. In the single-label folders, each folder is named according to a single class label. In the multi-label folders, the folder names consist of all class labels representing the labels of the objects contained within, separated by spaces.

The multi-label classifier requires the data converted from this folder structure into a single folder with a specification file defining name, label and validation set. This is achieved through a Python script which also randomly choose 20% of objects from each class as validation data. An example of the comma-separated values (CSV) specification file is shown in Table 1. The entire data collection process is summarized in Figure 3.

Table 1: CSV specification file format for multi-label classification. 20% validation (validation column “yes”).

fname	label	validation
object1.png	pump	no
object2.png	valve	yes
object3.png	valve	no

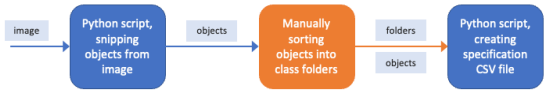


Figure 3: Data collection and preparation for training image classification models.

Data for object detection is attained through image annotation and will be a resulting part of the developed semi-automated annotation tool.

3.2. Image classification

In this project, the primary focus for the application of single-label classification is to assess the performance improvement of a pretrained ResNet50 model. The objective is to evaluate the effectiveness of transfer learning with custom data in comparison to training a fresh model. For more information on this topic, please refer to the Master’s thesis [9]. The single-label classifiers used in this study were trained on approximately 1000 object images.

In the future task of classifying objects within a region of interest (RoI), the utilization of multi-label classification with the ResNet50 model proves to be advantageous. This approach enables the classifier to detect multiple objects within the RoI, while also taking into account situations where no objects are present, thus minimizing the occurrence of misclassifications. The multi-label classifier is trained on approximately 1400 object images, where 1000 of these are single-label object images. It is important to ensure that for all multi-label object classes, there exist good representative

single-label object classes. For instance, if a multi-label object image contains both a valve and a line, it is necessary to have separate single-label object images for both the valve and the line classes. This approach enables the model to effectively distinguish between different objects and enhance its classification capabilities.

3.3. Semi-automated annotation

Combining three computer vision techniques: sliding window, image pyramid scaling, and Non-Maximum Suppression (NMS), to extract objects from an image. The sliding window divides the image into overlapping windows, generating potential RoI's. The image pyramid scaling ensures object detection at different scales. NMS eliminates redundant detections, selecting the most accurate bounding boxes. Extracted image snippets are classified using a multi-label classification model derived in the previous step. This approach serves as the foundation for the pre-analysis stage in the development of a semi-automated annotation tool.

The requirements and functionality desired for this tool include the ability to upload images, perform preprocessing, conduct pre-analysis using the above-mentioned method, manually adjust the pre-analysis annotations, and export the annotated data in a format suitable for one-stage or two-stage object detectors. Specifically, the tool should provide the option to export the annotations in a text file format compatible with the YOLO detector.

3.4. YOLOv8 object detector

The annotation process for generating training and validation data for modern object detectors has now been optimized. It is well-established that one-stage detectors are faster but often yield lower accuracy compared to two-stage detectors [20]. At the time of undertaking this project, a new one-stage detector for YOLO was introduced, promising improved mAP scores and better results on tiny objects. Considering the goal of analyzing a large volume of images in a single run, speed is crucial. Consequently, the latest YOLOv8 architecture provided by Ultralytics [21] have been chosen.

Ultralytics provides different sizes of their model. As a general guideline, larger models are capable of capturing more features than smaller ones [22-23]. After checking this guideline by evaluating the medium, large, and extra-large models using the same dataset, the largest model was selected. This testing process is listed in Table 2.

Table 2: Model sizes tested using custom dataset of 21 train and 3 validation images. 1000 epochs of training.

Model	Early stopping	mAP^{50}
medium	270 epochs	80.8%
large	388 epochs	83.0%
xlarge	326 epochs	90.5%

3.5. Extracting tags

OCR involves preprocessing the image, localizing text, character segmentation and recognition, and post-processing. The focus is on utilizing the recognition part of the Pytesseract OCR library, excluding post-processing dictionary translation. The tags in operator interface images consist of combinations of numbers, letters, and symbols. To filter out unwanted combinations, a custom dictionary is created. The recognized tags and their positions relative to the image are stored in a text file, with positions normalized between 0 and 1. Initially, OCR on the raw image did not provide valuable information due to small tag sizes, random placement, and low contrast. To address this, the image is divided/split into sections, a scale pyramid is applied, and preprocessing steps such as grayscale conversion, blurring, edge detection, and dilation are performed as shown in Figure 4. This significantly improves the OCR results as shown in Table 3.

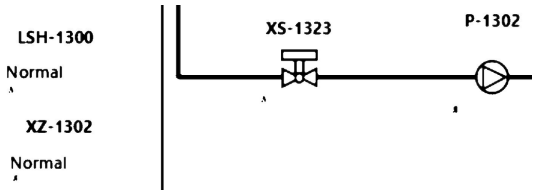


Figure 4: Preprocessing images to only identify tags and high contrast lines.

Table 3: Showing a small part of OCR result before and after preprocessing the image.

No image preprocessing	Custom image preprocessing
2341 *0.090365 0.093981 0.003385 0.002315*	***LSH-1300*** *0.089461 0.689971 0.030979 0.020058*
0201 *0.533594 0.083912 0.004687 0.002546*	***XS-1323*** *0.183917 0.687645 0.025043 0.020058*
0201 *0.611393 0.083796 0.003255 0.001852*	***P-1302*** *0.258107 0.683430 0.019687 0.023256*

3.6. Linking objects and tags

The concept behind tag extraction in conjunction with object detection is to establish a relationship between tags and objects based on their respective locations. It is reasonable to assume that tags and objects located close to each other are associated. However, there are certain arguments against this generalization. For instance, tags may be situated far from an object due to unobservable status variables associated with the object in the current image state. This distance may even exceed the distance between the tag and an unrelated object. Consequently, in this scenario, a single tag can be linked to multiple objects.

As demonstrated in the previous step, the OCR engine extracts texts and converts their positions to a normalized scale ranging from 0 to 1, matching the YOLOv8 object detection location scale. To determine the distance between the center of the tag location and the object location, the Minimum

Euclidean Distance calculation is employed. Visual representation of the distance calculation is shown in Figure 5.

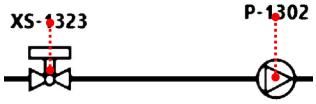


Figure 5: Visualized linking of object and tags.

3.7. ICE - Industrial Component Extraction tool

An analysis software that encompasses all the components, including the object detector, OCR and minimum Euclidean distance calculator, is created. Additionally, the software will provide a defined export format that allows users to easily view the final analysis. The final solution is structured as shown in the use case diagram in Figure 6.

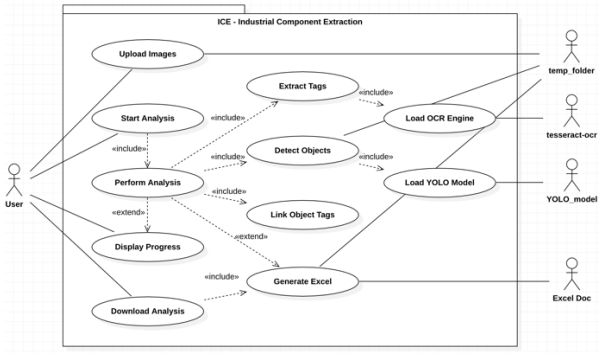


Figure 6: Use Case Diagram for final analysis software ICE – Industrial Component Extraction.

4. Results and Discussion

4.1. Multi-label image classification

A learner is defined, with minor data augmentation such as vertical and horizontal flipping, as well as zero-padding. Since the data never will be warped, and all information within the image is relevant, no further augmentation is needed. Figure 7 displays a sample from a training data minibatch.

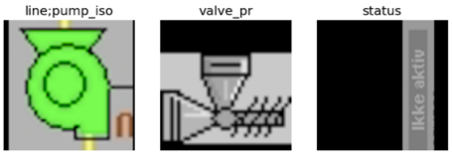


Figure 7: Sample of a minibatch.

When dealing with multi-label image classification, it is crucial to establish an appropriate multi-accuracy threshold. A threshold value of 0.8 is selected which is heigh and within a smoothness of FastAI's [24] threshold finders' curve for this function thus ensuring no outliers are selected [25].

After finetuning the ResNet50 model for 11 epochs (4 freeze and 7 un-freeze), an accuracy of 99.33% is achieved. By analyzing the loss plot shown in Figure 8 both the validation- and training-loss flattens. There is no indication of overfitting, so a score of 99.33% is acceptable and training is stopped after the 11th epoch.

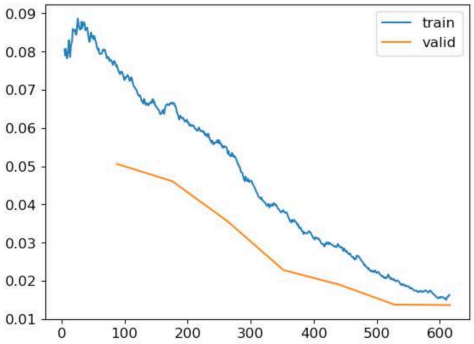


Figure 8: Multi-label classification loss plot.

4.2. Semi-automated annotation tool

The multi-label classifier in combination with pyramid scaling, sliding window and soft NMS results in a multi-class object detector. A customized version of the soft NMS is required to only allow same type labels to suppress each other, shown in Equation (1).

$$J_{label}(A_{label}, B_{label}) = \frac{|A_{label} \cap B_{label}|}{|A_{label} \cup B_{label}|} \quad (1)$$

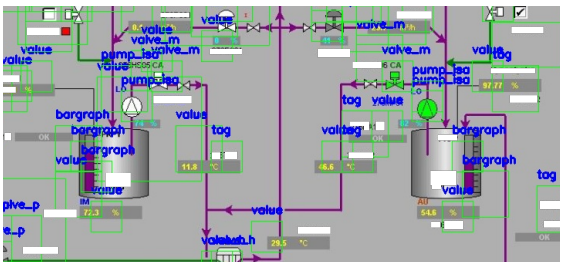


Figure 9: Small extract (snippet) of the multi-class object detector result.

From the image displayed in Figure 9, this method appears sub-optimal as it scores poorly on both position and classification. The resulting custom mAP^{50} calculation score of 5 individual images was 11.36%. More on how this custom accuracy calculation was performed in Master's thesis [9]. This method provides a reference to the object and its location, serving the purpose as a pre-analysis step for the semi-automated annotation tool. The goal is to enhance efficiency in annotating data for modern object detectors.

The annotation tool is developed as specified by the requirements, where a user can upload an image, pre-analyze it using the multi-class object detector with the multi-label classifier model trained in chapter 4.1, and further make modifications and improve the annotation. The user can now export a state-of-the-art annotation file for the object detector. It is estimated an 75% increased efficiency using this tool compared to traditional third-party tool due to the pre-analysis which

provides classes and a starting point for the user to annotate.

The tool was later modified to also take pre-annotated images, if existing. Which would give a good starting point of annotation (no pre-analysis needed by multi-class object detector).

4.3. YOLOv8 object detector

To prevent aliasing due to downscaling when loading the training and validation datasets to the object detector, the image and annotation data was split with a custom script. Thus, increasing the training data. This resulted in a better model.

At the final iteration, the YOLOv8 model was transfer learned on 59 training and 11 validation images from three different sites to improve generalization. An overview of the training iterations can be seen in Table 4.

Table 4: Iterations of training and validating the YOLOv8 transfer learned model. All runs are performed with parameters: patience=150, batch=8, model=xlarge.

Runs	Dataset	N Sites	mAP^{50}	Note
1	21 train, 3 val	1	90.5%	Tag classes included
2	21 train, 3 val	1	87.1%	Removed tag classes, fixed some errors
3	40 train, 8 val	1	97.2%	Realized non-generalized model
4	59 train, 11 val	3	95.5%	Added more data from different sites

Since the object detector returns an annotation file during testing on new data, sub-optimal tests that detect only a small percentage of objects can be fed back to the semi-automated annotation tool for improvement. Consequently, new test images are transformed into training and validation images, necessitating the need for more data. This created the idea of modifying the annotation tool to include the YOLOv8 object detector as a pre-analysis step instead of the multi-class object detector.

In summary, the misclassifications of the YOLOv8 model shown to the left in Figure 10 is strongly related to number of representatives in the dataset shown to the right in Figure 10.

4.4. ICE - Industrial Component Extraction tool

Combining the custom YOLOv8 object detector model with OCR and linking objects and tags leads to the development of effective analysis software. The software is designed based on the requirements and use case diagram depicted in Figure 6. The resulting software exhibits an estimated improvement in efficiency, approximately 10 times greater than manual analysis.

The Pytesseract OCR engine only extracts approximately 50% of tags due to non-customized model (depending on image to text size ratio). OCR image preprocessing is also the most time-consuming part of the analysis. Analyzing 12 images takes 1 minute and 20 seconds, where the object detection only uses approximately 25ms (on average) for each image. Image preprocessing and OCR account for most of the remaining time required.

The analysis results are exported in an Excel document format, which includes a summary sheet as the first page, providing an overview of the analyzed images and the detected objects in each image. Additionally, separate sheets are generated for each image, displaying the bounding box objects with labels within the images, shown in Figure 11. Moreover, a data sheet for each image is

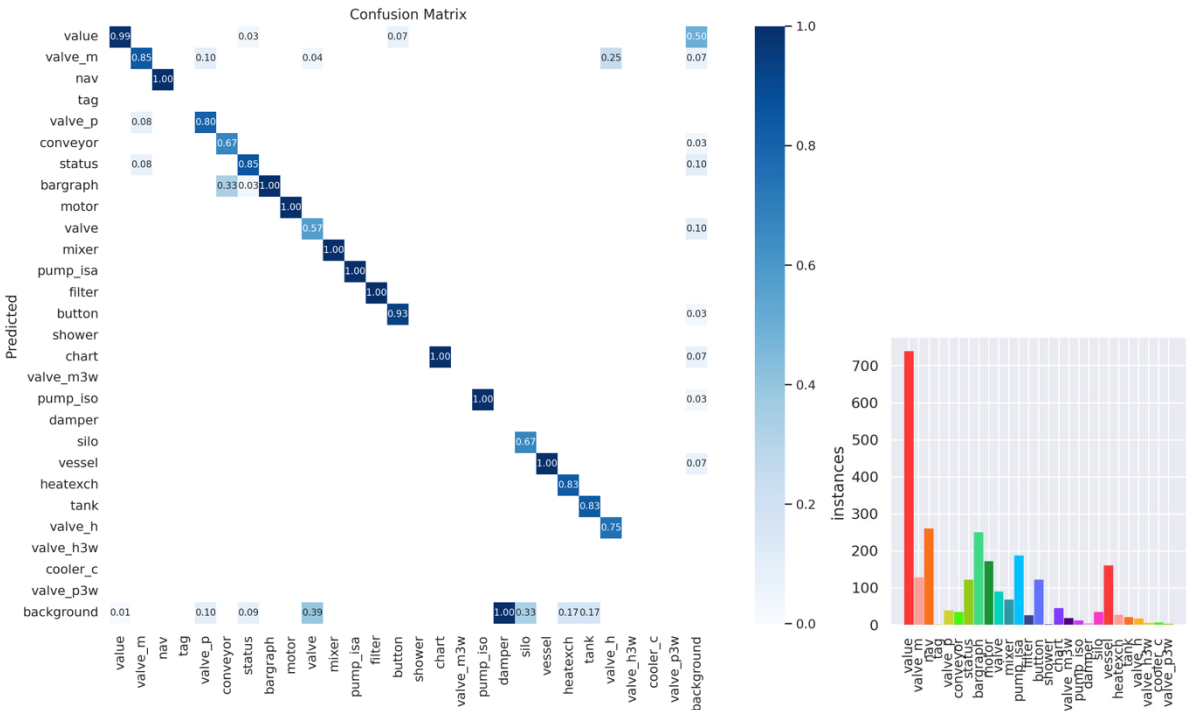


Figure 10: YOLOv8 confusion matrix (left) and number of instances (right), run 4.

included, showing each detected object snipping along with its associated tag and information in a table format, shown in Figure 12.

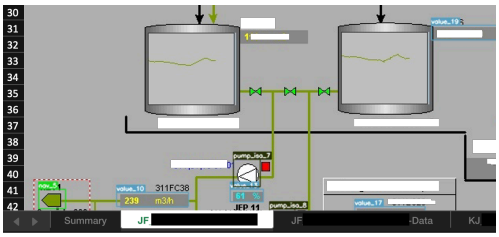


Figure 11: Individual image analysis sheet with bounding boxes.

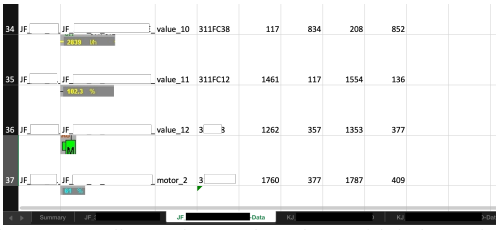


Figure 12: Adjacent image data sheet with information.

5. Future Work

5.1. Improvements to current solutions

The OCR tag extraction method used in this project can be improved by exploring alternative extraction methods or adding tags as a separate label class in the object detector. Treating tags as a separate class would require more training data but could improve text extraction by individually feeding tag objects to the OCR engine. The poor performance in the current solution is due to the small size and random placement of text compared to other objects. Multiple image scales and preprocessing were used to improve detection, but this increases computing power and is time consuming.

To enhance the final ICE software, incorporating features from the annotation tool would be beneficial. Users could perform pre-analysis on a portion of customer images using the YOLOv8 model, make manual adjustments to the detection, and retrain the model. Then upload the rest of the images and get an improved analysis with the improved model. This iterative process improves model generalization and user experience. It would be valuable to have these features available for all users of the ICE software for daily image analysis.

Including more annotated images from various sites would enhance the model's performance and generalization. Alternatively, modifying the YOLOv8 network architecture by substituting the classification network with the multi-label image classification model could eliminate the need for annotating more images.

5.2. Opportunities for future development

Combining the developed product with pixel processing for pipeline detections, as discussed by Moon et al. [8], could provide a solution for documenting structured image flows. Further training a large language model (LLM) on the source-code libraries for operator interface image designs would enable features for mapping of detected objects to generate prompts that results in automatically generating code for new images. Extending this concept to configuration and documentation such as system control diagrams (SCDs), and P&IDs, by training the LLM on relevant data, and improving models for detection and text extraction, could automate the process of redesign/migrating operator interface images entirely.

A solid object detection model for process graphics can also monitor real-time system images and extract information without direct system logic interaction. For instance, it can be useful in situations where integration with communication protocols is not feasible. As an example, placing a web camera in front of old HMI panels and extracting data to a cloud solution offers a solution when integration or modernization is not an option.

These advancements hold potential for diverse applications, improving efficiency in challenging environments.

6. Conclusion

In this project, an object detection system was developed for operator interface images, with a focus on optimizing data acquisition, annotation, model training, and software integration. The approach involved a semi-automated annotation tool that utilized multi-label classification and traditional computer vision techniques, resulting in an estimated 75% efficiency improvement compared to traditional tools. The tool supports both two-stage and one-stage detectors, allowing manual adjustments to analysis and exporting annotations in popular formats.

Next, the utilization of the YOLOv8 model from Ultralytics was explored, and it was trained with custom data generated using the semi-automated annotation tool. After multiple iterations and preprocessing techniques on 70 images, a mAP^{50} score of 95.5% was achieved. The final model was then integrated into a user-friendly web application that enables users to upload images, perform analyses, and obtain downloadable results in an Excel format. This tool streamlines project planning, improves efficiency, and facilitates cost estimation for migration projects. Estimating a reduction of time spent analyzing HMI by 90% compared to the manual approach.

The project successfully established a novel foundation for object detection in operator interface images, providing an efficient semi-automated annotation tool and a high-performing YOLOv8 model. The developed software application has the

potential to enhance project planning efficiency and accuracy, benefiting various industries. However, it is essential to emphasize the need for thorough data collection and testing to ensure the accuracy and generalizability of the model.

Acknowledgment
Thanks to industrial partner and employer Emerson Automation Solutions.

References

[1] C. C. Editor, "Human-Machine Interface (HMI) - Glossary | CSRC." https://csrc.nist.gov/glossary/term/human_machine_interface (accessed Jun. 04, 2023).
[2] Inductive Automation, "What is SCADA?," *Inductive Automation*. http://www.inductiveautomation.com/resources/article/what-is-scada (accessed May 05, 2023).
[3] K. Balaji and K. Lavanya, "Image Classification - an overview | ScienceDirect Topics." https://www.sciencedirect.com/topics/engineering/image-classification (accessed May 08, 2023).
[4] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object Detection in 20 Years: A Survey." arXiv, Jan. 18, 2023. Accessed: Mar. 23, 2023. [Online]. Available: http://arxiv.org/abs/1905.05055
[5] S. Paliwal, M. Sharma, and L. Vig, *OSSR-PID: One-Shot Symbol Recognition in P&ID Sheets using Path Sampling and GCN*. 2021.
[6] J. Deng, J. Guo, J. Yang, N. Xue, I. Kotsia, and S. Zafeiriou, "ArcFace: Additive Angular Margin Loss for Deep Face Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 5962–5979, Oct. 2022, doi: 10.1109/TPAMI.2021.3087709.
[7] R. Rahul, S. Paliwal, M. Sharma, and L. Vig, "Automatic Information Extraction from Piping and Instrumentation Diagrams:," in *Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods*, Prague, Czech Republic: SCITEPRESS - Science and Technology Publications, 2019, pp. 163–172. doi: 10.5220/0007376401630172.
[8] Y. Moon, J. Lee, D. Mun, and S. Lim, "Deep Learning-Based Method to Recognize Line Objects and Flow Arrows from Image-Format Piping and Instrumentation Diagrams for Digitization," *Applied Sciences*, vol. 11, no. 21, Art. no. 21, Jan. 2021, doi: 10.3390/app112110054.
[9] E. Illing, "Object detection, information extraction and analysis of operator interface images using computer vision and machine learning.," Masters Thesis, University of South-Eastern Norway, Porsgrunn, Norway, 2023.
[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition." arXiv, Dec. 10, 2015. Accessed: Mar. 29, 2023. [Online]. Available: http://arxiv.org/abs/1512.03385
[11] V. Wiley and T. Lucas, "Computer Vision and Image Processing: A Paper Review," *International Journal of Artificial Intelligence Research*, vol. 2, p. 22, Feb. 2018, doi: 10.29099/ijair.v2i1.42.
[12] Nichole Peterson, "History of Computer Vision and Its Principles | alwaysAI Blog | alwaysAI." https://www.alwaysai.co/blog/history-computer-vision-principles (accessed Mar. 23, 2023).
[13] IBM, "What is Data Labeling? | IBM." https://www.ibm.com/topics/data-labeling (accessed Jun. 04, 2023).
[14] Jeremy Jordan, "An overview of object detection: one-stage methods.," *Jeremy Jordan*, Jul. 11, 2018. https://www.jeremyjordan.me/object-detection-one-stage/ (accessed Jun. 04, 2023).
[15] Sovit Rath, "YOLOv8 Ultralytics: State-of-the-Art YOLO Models," Jan. 10, 2023. https://learnopencv.com/ultralytics-yolov8/ (accessed Mar. 24, 2023).
[16] J. Solawetz, F. JAN 11, and 2023 10 Min Read, "What is YOLOv8? The Ultimate Guide.," *Roboflow Blog*, Jan. 11, 2023. https://blog.roboflow.com/whats-new-in-yolov8/ (accessed Mar. 24, 2023).
[17] Zelic Filip and Anuj Sable, "Tesseract OCR in Python with Pytesseract & OpenCV," *Nanonets AI & Machine Learning Blog*, Aug. 09, 2022. https://nanonets.com/blog/ocr-with-tesseract/ (accessed Feb. 23, 2023).
[18] exposit_marketing, "Computer Vision Object Detection: challenges faced," *Exposit*, Apr. 20, 2021. https://www.exposit.com/blog/computer-vision-object-detection-challenges-faced/ (accessed Jun. 04, 2023).
[19] T. OpenCV, "About," *OpenCV*. https://opencv.org/about/ (accessed Feb. 23, 2023).
[20] X. Lu, Q. Li, B. Li, and J. Yan, "MimicDet: Bridging the Gap Between One-Stage and Two-Stage Object Detection," in *Computer Vision – ECCV 2020*, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds., in Lecture Notes in Computer Science, vol. 12359. Cham: Springer International Publishing, 2020, pp. 541–557. doi: 10.1007/978-3-030-58568-6_32.
[21] Ultralytics, "Quickstart - YOLOv8 Docs." https://docs.ultralytics.com/quickstart/ (accessed Mar. 24, 2023).
[22] A. Kumar, "Model Complexity & Overfitting in Machine Learning," *Data Analytics*, May 29, 2022. https://vitalflux.com/model-complexity-overfitting-in-machine-learning/ (accessed Jun. 04, 2023).
[23] T. FastAI, "FastAI 05_pret_breeds." https://colab.research.google.com/github/fastai/fastbook/blob/master/05_pet_breeds.ipynb (accessed May 08, 2023).
[24] J. H. and S. Gugger, "fast.ai - fastai A Layered API for Deep Learning." https://www.fast.ai/posts/2020-02-13-fastai-A-Layered-API-for-Deep-Learning.html (accessed Feb. 23, 2023).
[25] T. FastAI, "FastAI 06_multicat." https://colab.research.google.com/github/fastai/fastbook/blob/master/06_multicat.ipynb#scrollTo=invs-Qyn8lSC (accessed Feb. 27, 2023).

Modeling and identification of the Quanser Aero using a detailed description of friction and centripetal forces

Mathias Dyvik[1] Didrik Efjestad Fjereide[1] Damiano Rotondo*[1]

[1]Department of Electrical and Computer Engineering (IDE), University of Stavanger, Stavanger, Norway
Corresponding author: damiano.rotondo@uis.no

Abstract

This paper deals with the modeling and identification of the Quanser Aero. The Quanser Aero is an aerospace laboratory setup designed for teaching aerospace concepts. Two propellers generate thrust and allow the user to control its dynamic response. The ability to lock axes individually makes it capable of abstracting a variety of aerospace systems, such as half-quadrotor, 1-Degree of freedom (DOF), vertical take-off and landing (VTOL), and 2-DOF helicopter. This paper focuses on the latter of these modes. In this configuration, the Quanser Aero can produce different pitch and yaw angles based on the angular velocity of the propellers, which produces an interesting identification and control problem, due to the presence of nonlinearities and significant cross-couplings between different variables. In this paper, a nonlinear model derived from Newton's law and Euler's rotational dynamics is obtained, and the unknown model parameters are identified through an experimental approach, with the model validated through real-time testing. In particular, it is shown that by means of a more detailed description of the friction, which includes the Karnopp's model that keeps the sum of the magnitude of all forces equal to zero until the applied forces are strong enough to overcome the friction force, and of the centripetal forces acting on the Aero, significant improvements are obtained when compared to state-of-the-art models. These improvements may hold the potential to enhance the performance of advanced nonlinear model-based control algorithms for this device.
Keywords: Quanser Aero, nonlinear model, friction

1 Introduction

With the advance of technology, aerial vehicles are steadily becoming more and more ubiquitous. For example, their unmanned versions, commonly referred to as *drones*, are widely used in several applications where it is either dangerous or economically inefficient to employ human pilots, e.g., for monitoring (Ren et al., 2019) and transport (Thiels et al., 2015) tasks. In order to train engineers to analyse the dynamical behaviour of aerial vehicles and design efficient and effective feedback control systems for them, small-scale devices that are apt to use in laboratories for educational purposes have been developed. These devices solve some of the most common issues encountered when unmanned aerial vehicles are applied to education, such as their limited autonomy and logistical/safety problems, among others (Shadiev and Yi, 2022). One of such devices is the Quanser Aero, a reconfigurable dual-rotor aerospace setup designed for teaching control and introducing aerospace concepts at an undergraduate level, but at the same time complex enough to enable its use in research, see e.g. (Kim and Ahn, 2021; Fliess and Join, 2022).

Attaining a satisfactory performance when controlling a highly coupled and nonlinear system such as the Quanser Aero is a challenging problem. It is known that nonlinear control algorithms are able to achieve better performance than linear control schemes (Kumar and Dewan, 2022), although they require a precise and reliable model that describes the system under consideration over a broad range of operating conditions. For this reason, several students and researchers have engaged in obtaining a model of the Quanser Aero that goes beyond the simple linear model provided by the manufacturer. For instance, (Frasik and Gabrielsen, 2018) have provided a nonlinear model calibrated through the analysis of free-oscillation and step responses. (Schlanbusch, 2019) derived a nonlinear mathematical model based on an Euler-Lagrange approach, where the uncertain parameters were updated in real-time by an adaptive control law.(Abdelwahed et al., 2019) used a reduced complexity autoregressive with exogenous input (ARX)-Laguerre model to describe the Quanser Aero in a selected operating point. (Segerstrom et al., 2021) used particle swarm optimization and constrained nonlinear optimization to obtain key electrical, mechanical and aerodynamical parameters for this device. (Kumar and Dewan, 2022) further improved the model and used it to design linear quadratic regulator (LQR) and sliding mode controller (SMC) algorithms.

In this paper, we obtain a nonlinear model using a Newton-Euler approach, identify its unknown parameters through an experimental approach, and validate the model through real-time testing. The main feature of the proposed model is a more detailed description of the friction, which includes the Karnopp's model, and of the centripetal forces acting on the Aero. Using different per-

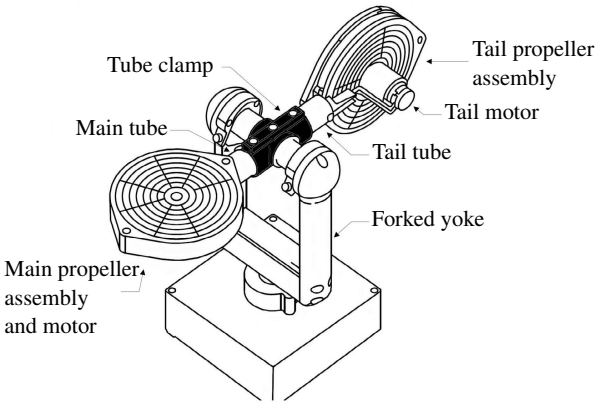


Figure 1. General component arrangement of the mechanical parts on the Quanser Aero. Adapted from (Qua, 2016)

formance indices under several operating conditions, we highlight significant improvements in the model's ability to capture the nonlinearities of the system in comparison with state-of-the-art models. Hence, the proposed model could potentially contribute significantly to the design of advanced nonlinear model-based control algorithms that require a stronger match between reality and model.

The remaining of this paper is structured as follows. Section 2 provides a description of the Quanser Aero. A detailed explanation of the mathematical model considered in this paper is given in Section 3. Section 4 presents the experimental comparison between the proposed model and other state-of-the-art models encountered in the literature. Finally, Section 5 summarizes the main conclusions of the work.

2 System description

The Aero is a dual-propeller laboratory setup designed for education and research in aerospace control, manufactured by Quanser Consulting Inc. It is equipped with two identical propellers, each of which is driven by a brushed direct-current (DC) motor. By applying motor voltages ranging within ± 18V, the user can manipulate the thrust produced by the propellers, thereby affecting the dynamics of the Aero. Moreover, the experiment is reconfigurable, meaning that the propeller assemblies can be tilted to desired angles. This, together with its ability to lock axes individually, allows the Aero to abstract various aerospace systems, such as half-quadrotor 1-DOF VTOL, and 2-DOF helicopter. This paper considers the latter case, where the propellers are perpendicular, as shown in Fig. 1.

The Quanser Aero can be considered as the composition of three structures: the base, the support yoke, and the helicopter body. The base is a stationary box at the bottom of the Aero and contains the electrical and electronic components necessary to control the system. The support yoke, shaped like a fork, stands vertically on the Aero base and serves to elevate the helicopter body, allowing it to rotate around its vertical axis. A rotational joint between the base and the yoke combined with a slip ring wiring system allows for an unlimited 360° yaw rotation. The helicopter body comprises two coupled metal tubes held together by a tube clamp, with a pair of propeller assemblies at both ends. It is horizontally attached to the forked yoke through a rotational joint, allowing the pitch angle of the helicopter body to rotate between -62° and $+54^\circ$.

Figure 2. 3D-printed low-efficiency (left) and high-efficiency (right) propellers (left)

The propeller pair in the Quanser Aero system is interchangeable. As of now, Quanser offers two types of propellers: high-efficiency and low-efficiency (see Fig. 2). This study considers the latter type. Although the high-efficiency propellers are more representative of small-scale aerial vehicles, a dynamic coupling between axes, which is seen in real aerial vehicles, is only present when the low-efficiency propellers are used. Thus, from a control engineering perspective, the use of low-efficiency propellers poses a more interesting and challenging problem.

The Aero setup is equipped with four optical rotary encoders that precisely measure the angular position of each DC motor, as well as the pitch and yaw angles of the helicopter body. An inertial measurement unit (IMU), which combines an accelerometer and a gyroscope, is also integrated into the Aero.

3 Mathematical Model

The mathematical model presented in this section describes the 2-DOF helicopter configuration of the Quanser Aero. It was derived by considering it as a rigid body and applying Newton's laws and Euler's rotational dynamics. The model is based on several electrical and mechanical parameters outlined in Table 1, most of which are provided by the manufacturer. Unknown parameters related to dynamical properties such as damping, friction, and propeller thrust have been experimentally identified and are listed in Table 2.

3.1 Propeller dynamics

The main and tail propeller dynamics are modeled as two nonlinear differential equations containing identical parameters:

$$\frac{\mathrm{d}\omega_p}{\mathrm{d}t} = \frac{K_\tau}{R_a J_{eq}} v_p - \frac{K_\tau K_E}{R_a J_{eq}} \omega_p - \frac{1}{J_{eq}} f_d(\omega_p) \tag{1}$$

Table 1. Known/ estimated physical parameters

Symbol	description	Value
Dc motor		
K_τ	Torque constant	$0.042\ Nm/a$
K_E	Motor back-emf constant	$0.042\ V/rad \cdot s$
R_a	Terminal resistance	$8.4\ \Omega$
J_{rotor}	Rotor inertia	$4.0 \times 10^{-6}\ kgm^2$
m_e	Mass of DC motor	$0.200\ kg$
Propeller		
J_{prop}	Low-efficiency propeller inertia	$3.2 \times 10^{-5}\ kgm^2$
J_{hub}	Propeller hub inertia	$3.04 \times 10^{-9}\ kgm^2$
m_{pa}	Mass of propeller assembly	$0.146\ kg$
Helicopter body		
m_b	Total mass of helicopter body	$1.15\ kg$
m_{mt}	Mass of main propeller tube	$0.089\ kg$
m_{tt}	Mass of tail propeller tube	$0.089\ kg$
m_{tc}	Mass of tube clamp	$0.280\ kg$
l_t	Length of tubes and tube clamp when assembled	$0.165\ m$
d_t	Thrust displacement	$0.158\ m$
d_m	Aero body vertical COM displacement	$0.00325\ m$
d_c	Main/tail section horizontal COM displacement	$0.106\ m$
Forked yoke		
m_y	Mass of the forked yoke	$0.526\ kg$
r_y	Radius of the forked yoke	$0.02\ m$

$$\frac{\mathrm{d}\omega_y}{\mathrm{d}t} = \frac{K_\tau}{R_a J_{eq}} v_y - \frac{K_\tau K_E}{R_a J_{eq}} \omega_y - \frac{1}{J_{eq}} f_d(\omega_y) \tag{2}$$

where $v_{p/y}$ is the input voltage of the main/tail motor, $\omega_{p/y}$ is the angular velocity of the main/tail propeller, and J_{eq} is the total moment of inertia of the main/tail propeller subsystem.

The torques exerted by the DC motors on the propeller shafts can be modelled as the product of a torque constant, K_τ, and the applied armature currents, $i_{ap/y}$:

$$\tau_{mp/y} = K_\tau i_{ap/y} \tag{3}$$

The armature current can then be expressed in terms of the input voltages by applying Kirchhoff's voltage law:

$$i_{ap/y} = \frac{v_{p/y}}{R_a} - \frac{L_a}{R_a} \frac{di_{ap/y}}{dt} - \frac{e_b}{R_a} \tag{4}$$

where R_a is the motor armature resistance, L_a is the motor inductance, and $e_b = K_E \omega_{p/y}$ is the back-emf voltage which depends on the angular velocity of the propeller shafts. As suggested by the Quanser laboratory guide, (4) can be simplified by neglecting electrical dynamics, i.e., setting the motor inductance equal to zero (Qua, 2016). Furthermore, the Quanser lab guide proposes modeling the torque exerted by drag and air resistance as linear functions of propeller angular velocities:

$$\tau_d = k_d \omega_m \tag{5}$$

where k_d is an experimentally derived torque constant. However, online experiments conducted on the Aero's main propeller have shown that the following non-linear function provides a more accurate representation of the drag and air resistance:

$$\begin{aligned} \tau_d = f_d(\omega_{p/y}) = sign(\omega_{y/p}) k_{d1} \omega_{p/y}^2 \\ + k_{d2} \omega_{p/y} + sign(\omega_{y/p}) k_{d3} \end{aligned} \tag{6}$$

Table 2. Experimentally identified parameters

Parameter	Value	Parameter	Value
k_{Mpp_1}	1.69×10^{-6}	k_{Typ_1}	1.06×10^{-6}
k_{Mpp_2}	9.65×10^{-7}	k_{Typ_2}	1.17×10^{-5}
k_{Mpn_1}	-2.55×10^{-6}	k_{Tyn_1}	-1.41×10^{-6}
k_{Mpn_2}	-4.69×10^{-5}	k_{Tyn_2}	-4.16×10^{-5}
k_{Myp_1}	7.30×10^{-7}	k_{Tpp_1}	1.66×10^{-6}
k_{Myp_2}	-1.61×10^{-5}	k_{Tpp_2}	-2.83×10^{-5}
k_{Myn_1}	-6.43×10^{-7}	k_{Tpn_1}	-1.63×10^{-7}
k_{Myn_2}	3.28×10^{-5}	k_{Tpn_2}	8.37×10^{-6}
k_{d_1}	2.90×10^{-7}	k_{Dyp_1}	1.84×10^{-5}
k_{d_2}	4.20×10^{-6}	k_{Dyp_2}	3.64×10^{-4}
k_{d_3}	8.00×10^{-4}	k_{Dyn_1}	-5.05×10^{-5}
k_{Fyp}	4.98×10^{-3}	k_{Dyn_2}	9.86×10^{-4}
k_{Fyn}	-2.90×10^{-3}	k_{Dp_1}	7.10×10^{-3}
k_{Fp}	2.00×10^{-4}		

where k_{d1}, k_{d2}, and k_{d3} are torque constants identified by iterative online tuning.

The total moment of inertia, J_{eq}, has been modeled according to the Quanser laboratory documentation as:

$$J_{eq} = J_{prop} + J_{hub} + J_{rotor} \tag{7}$$

where J_{prop} denotes the moment of inertia of the propeller, J_{hub} represents the moment of inertia acting on the clamp used to mount the propeller to the motor, and J_{rotor} is the moment of inertia of the DC motor shaft. The specific values for J_{hub} and J_{rotor} can be obtained from the Quanser laboratory documentation. However, the moment of inertia for the propeller, J_{prop}, is only provided for the high-efficiency propeller. The moment of inertia for the low-efficiency propeller has therefore been estimated using the computer-aided design (CAD) software *Autodesk Inventor*.

The resulting propeller velocity model (1)-(2) is obtained by considering equations (3), (6) and (7), and applying Newton's second law for rotation: $J\ddot{\theta} = \sum_i \tau_i$.

3.2 Pitch motion

The model describing the Aero's pitch movement is expressed as:

$$\begin{aligned} \frac{\mathrm{d}\Omega_p}{\mathrm{d}t} = & \frac{f_{Mp}(\omega_p) d_t + f_{Tp}(\omega_y) d_t - m_b g d_m \sin(\theta_p)}{J_p} \\ & - \frac{(m_A + m_B) \Omega_y^2 d_c^2 \cos(\theta_p) \sin(\theta_p)}{J_p} \\ & - \frac{f_{Dp}(\Omega_p) - f_{Fp}(\Omega_p)}{J_p} \end{aligned} \tag{8}$$

$$\frac{\mathrm{d}\theta_p}{\mathrm{d}t} = \Omega_p \tag{9}$$

where Ω_p is the angular velocity in the pitch direction, θ_p is the pitch angle and J_p is the moment of inertia about the pitch axis.

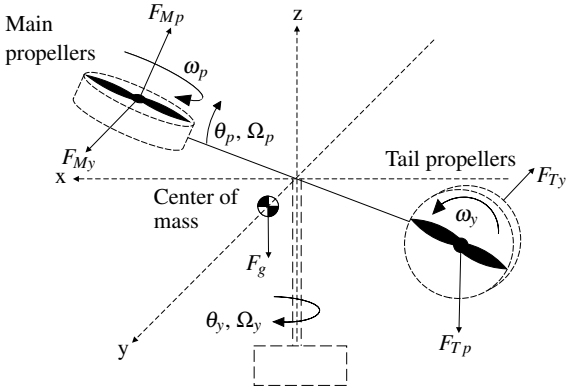


Figure 3. Free body diagram of the Quanser Aero 2-DOF helicopter configuration.

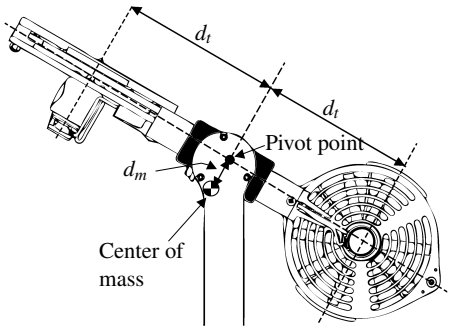


Figure 4. Center of mass and thrust displacement. Adapted from (Qua, 2016)

Fig. 3 shows a free-body diagram illustrating the forces acting on the Aero. In the 2-DOF helicopter configuration, the main propeller is fixed horizontally and produces a vertical thrust force F_{Mp} that enables the Aero to pitch. However, the main propeller also produces a perpendicular force F_{My}, which produces a cross-torque and thus a negative yaw rotation. The tail propeller is fixed vertically and is mainly used to counteract the cross-torque caused by the main propeller. The predominant thrust force from the tail propeller, which causes positive yaw rotation, is denoted by F_{Ty}. Similar to the main propeller, the tail propeller produces an additional perpendicular force F_{Tp}, which causes a positive rotation about the pitch axis. The gravitational pull F_g acts on the centre of mass (COM) of the helicopter body, which is slightly below its pivot point, so that the helicopter body acts like a pendulum. This implies that the pitch angle of the Aero returns to a horizontal equilibrium position if no force is applied. Fig. 4 illustrates the COM displacement d_m, which varies with how the propeller assemblies are angled. Unfortunately the documentation provided by Quanser Inc. gives the COM displacement only for the 1-DOF VTOL configuration. (Frasik and Gabrielsen, 2018) estimated this value to be $2.7mm$ for the 2-DOF helicopter configuration, whereas (Schlanbusch, 2019) suggest a higher value $3.8mm$. Our testing suggests that using the mean of the aforementioned values results in a satisfactory representation of the real system.

The torque induced by the gravitational force, which affects the pitch movement, is derived by multiplying the gravitational force F_g with the moment arm $d_m \sin(\theta_p)$:

$$\tau_g = F_g d_m \sin(\theta_p) = m_b g d_m \sin(\theta_p) \tag{10}$$

where m_b is the mass of the helicopter body.

The predominant thrust force of the main propeller F_{Mp} and the cross-thrust of the tail propeller F_{Tp} have been estimated using the experimental procedure proposed in (Schlanbusch, 2019). This method involves accelerating the propeller of interest to various rotational velocities while the other propeller remains disengaged, with the yaw axis locked, thus reducing the Aero setup to a 1-DOF system. The resulting steady-state pitch angles are then measured to obtain estimates of the thrust forces. With only the main/tail propeller engaged, the pitch angle at steady state, and the yaw axis locked, the equation of motion about the pitch axis with respect to $F_{Mp/Tp}$ becomes:

$$F_{Mp/Tp} = \frac{m_b d_m \sin(\theta_p)}{d_t} \tag{11}$$

where d_t is the thrust displacement of the Aero illustrated in Fig. 4. Then with the use of a least square approach, it was found that the following nonlinear piecewise functions provided a more accurate representation of the thrust forces:

$$F_{Mp} = f_{Mp}(\omega_p) = \begin{cases} k_{Mpp_1}\omega_p^2 + k_{Mpp_2}\omega_p & \text{if } \omega_p \geq 0 \\ k_{Mpn_1}\omega_p^2 + k_{Mpn_2}\omega_p & \text{if } \omega_p < 0 \end{cases} \tag{12}$$

$$F_{Tp} = f_{Tp}(\omega_y) = \begin{cases} k_{Tpp_1}\omega_y^2 + k_{Tpp_2}\omega_y & \text{if } \omega_y \geq 0 \\ k_{Tpn_1}\omega_y^2 + k_{Tpn_2}\omega_y & \text{if } \omega_y < 0 \end{cases} \tag{13}$$

The Coulomb friction opposing the pitch movement is modelled using a static friction coefficient:

$$F_{Fp} = f_{Fp}(\Omega_p) = sign(\Omega_p) k_{Fp} \tag{14}$$

while a linear function identified by Quanser (Qua, 2016) has been implemented for the pitch damping:

$$f_{Dp}(\Omega_p) = k_{Dp_1}\Omega_p \tag{15}$$

where k_{Dp_1} is a damping coefficient.

As the Aero undergoes a yaw rotation, the helicopter body's ends trace a spherical path. This implies that the main and tail sections will experience a centripetal force directed toward the centre of curvature, forcing the pitch position toward its horizontal equilibrium. The centripetal force acting on the main/tail part of the helicopter body can be expressed as:

$$F_{cm/t} = m_{A/B}\Omega_y^2 d_c \cos(\theta_p) \tag{16}$$

where $m_A = \frac{m_{tc}}{2} + m_{mt} + m_e + m_{pa}$ and $m_B = \frac{m_{tc}}{2} + m_{tt} + m_e + m_{pa}$, d_c is the length from the pivot point to the total

COM of the main/tail tube, main/tail propeller assembly, and half of the tube clamp. This distance has been calculated based on the dimensions of the tube clamp, which measures approximately 90mm, the length of each tube, which is around 130mm, and the distance from the pivot point to the centre of each propeller assembly, which is 158mm. The net torque produced by the vertical component of the centripetal forces acting on the tail and main sections of the Aero has been derived following the approach described in (Christensen et al., 2006), which involves a comprehensive analysis of a similar small-scale aerospace device. The resulting net centripetal torque is expressed as:

$$\tau_{Cp} = -(m_A + m_B)\Omega_y^2 d_c^2 \cos(\theta_p)\sin(\theta_p) \tag{17}$$

The resulting pitch motion model (8) is obtained by considering equations (10), (12)-(15), (17) and (31), and applying Newton's second law for rotation. Note that the torques produced by the thrust forces in (12) and (13) are computed by multiplying the forces by the moment arm d_t.

3.3 Yaw motion

The model describing the Aero's yaw movement is expressed as:

$$\frac{\mathrm{d}\Omega_y}{\mathrm{d}t} = \frac{f_{Ty}(\omega_y)\cos(\theta_p)d_t - f_{My}(\omega_p)\cos(\theta_p)d_t}{J_y(\theta_p)} - \frac{f_{Dy}(\Omega_y) - f_{Fy}(\Omega_y, F_{My}, F_{Ty})}{J_y(\theta_p)} \tag{18}$$

$$\frac{\mathrm{d}\theta_y}{\mathrm{d}t} = \Omega_y \tag{19}$$

where Ω_y is the angular velocity in the yaw direction, θ_y is the yaw angle, and $J_y(\theta_p)$ is the moment of inertia about the yaw axis.

The predominant thrust force of the tail propeller F_{Ty} and the cross-thrust of the main propeller F_{Mp} follow the same model as (12) and (13):

$$F_{My} = f_{My}(\omega_p) = \begin{cases} k_{Myp1}\omega_p^2 + k_{Myp2}\omega_p & \text{if } \omega_p \geq 0 \\ k_{Myn1}\omega_p^2 + k_{Myn2}\omega_p & \text{if } \omega_p < 0 \end{cases} \tag{20}$$

$$F_{Ty} = f_{Ty}(\omega_y) = \begin{cases} k_{Typ1}\omega_y^2 + k_{Typ2}\omega_y & \text{if } \omega_y \geq 0 \\ k_{Tyn1}\omega_y^2 + k_{Tyn2}\omega_y & \text{if } \omega_y < 0 \end{cases} \tag{21}$$

As the propellers are identical, it can be assumed that the parameters identified for f_{Mp} and f_{Tp} are representative for f_{Ty} and f_{My}, respectively. However, real-time testing indicated that this assumption resulted in more thrust than needed for the yaw movement due to unmodeled dynamics. Slightly down-tuned versions of the vertical thrust force functions are therefore implemented for the horizontal thrust force functions.

The viscous damping acting about the yaw axis, $f_{Dy}(\Omega_y)$ has been estimated by examining free-oscillation responses, as proposed in (Qua, 2016). To obtain the free-oscillation responses, the Aero was decelerated from various yaw velocities with both propellers disengaged and the pitch axis locked in its horizontal equilibrium position. By approximating the yaw deceleration as $\frac{\Delta\Omega_y}{\Delta t}$, (18) reduces to:

$$\frac{\Delta\Omega_y}{\Delta t} = -\frac{f_{Dy}(\Omega_y) + F_{Fy}}{J_y(0)} \tag{22}$$

where F_{Fy} is the static friction force opposing the yaw rotation. Then, by applying a least square regression to the measured data, a relationship between the angular yaw velocity and the sum of yaw damping and friction was obtained. The damping and friction have been distinguished by modeling the viscous damping using the regression coefficients and the static friction as the regression constant, as suggested in (Frasik and Gabrielsen, 2018), resulting in the following yaw damping function:

$$f_{Dy}(\Omega_y) = \begin{cases} k_{Dyp_1}\Omega_y^2 + k_{Dyp_2}\Omega_y & \text{if } \Omega_y \geq 0 \\ k_{Dyn_1}\Omega_y^2 + k_{Dyn_2}\Omega_y & \text{if } \Omega_y < 0 \end{cases} \tag{23}$$

and static friction force :

$$F_{Fy} = \begin{cases} k_{Fyp} & \text{if } \Omega_y > 0 \\ k_{Fyn} & \text{if } \Omega_y < 0 \end{cases} \tag{24}$$

However, during real-time testing of the Aero system, it was observed that the specific values obtained for the parameters $k_{Fyp/n}$ from the aforementioned experiment were too high. As a result, the parameter k_{Fyp} has been reduced by 20% and k_{Fyn} by 35%. Furthermore, to ensure that the net force acting on the yaw axis remains zero until the applied forces exceed the friction force, a friction modeling technique known as Karnopp's model has been implemented. In this way, the basic friction model in (24) can be extended to be valid at zero yaw velocity (Egeland and Gravdahl, 2002):

$$f_{Fy}(\Omega_y, F_{My}, F_{Ty}) = \begin{cases} sat(F_{Ty} - F_{My}, F_{Fy}) & \text{when } \Omega_y = 0 \\ F_{Fy} & \text{else} \end{cases} \tag{25}$$

where $sat()$ refers to the saturation function.

The resulting yaw motion model (18), is obtained by considering equations (20)-(21), (23), (25)-(26), and applying Newton's second law for rotation. Here, the torque produced by the thrust forces in (12) and (13) are computed by multiplying the thrust force by the moment arm $\cos(\theta_p)d_t$.

3.4 Moment of Inertia about the pitch and yaw axes

The total moments of inertia about the pitch/yaw axes J_p and J_y have been obtained as suggested in the laboratory documentation provided by Quanser Inc. However, by applying a more detailed description of the mass of the different mechanical components of the Aero, slightly lower numerical values for both the pitch and yaw inertia were obtained, which produced better results when the model

was tested against the actual plant. The components used in these estimates are shown in Fig. 1. Furthermore, Quanser's documentation ignores the fact that yaw inertia decreases with increasing pitch angle due to the mass being more concentrated along the vertical axis. Therefore, the yaw inertia is modelled as the nonlinear function of the pitch angle $J_y(\theta_p) = k_{Jy} f_{ky}(\theta_p)$, where k_{Jy} is the moment of inertia corresponding to a pitch angle of zero. In (Frasik and Gabrielsen, 2018) it was discovered that the nonlinear dynamics of the moment of inertia about the yaw axis are well described by a cosine function, resulting in the expression:

$$J_y(\theta_p) = k_{Jy}\cos(\theta_p) \tag{26}$$

The moments of inertia of the tubes and tube clamp connecting the main and tail propeller assemblies can be modelled as a solid cylinder perpendicular to the axis of rotation, rotating about its centre, resulting in:

$$J_{cylinder} = \frac{1}{12}(m_{mt} + m_{tt} + m_{tc})l_t^2 \tag{27}$$

By considering the two propeller assemblies as single-point masses rotating at a distance d_t from the pivot point of the helicopter body, their moment of inertia can be estimated as:

$$J_{pa} = (m_{pa} + m_e)d_t^2 \tag{28}$$

The forked yoke used to raise the helicopter body from the Aero base can be modelled as a solid cylinder rotating about its centre which leads to the expression:

$$J_{yoke} = \frac{1}{2}m_y r_y^2 \tag{29}$$

The total moment of inertia about the yaw axis with a fixed pitch angle of zero degrees is then obtained by summing (27), (28) and (29):

$$J_y(\theta_p = 0) = k_{Jy} = J_{cylinder} + 2J_{pa} + J_{yoke} \tag{30}$$

The total moment of inertia about the pitch axis is found as the sum of (27) and (28):

$$J_p = J_{cylinder} + 2J_{pa} \tag{31}$$

4 Experimental results

The purpose of the experiments described in this section was to test the model's ability to accurately predict the pitch angle, pitch velocity, yaw angle, and yaw velocity of the system when subjected to different input signals. To evaluate the performance of the model under various operating conditions, the input signal to the system was either varied in amplitude by changing the voltage level, or shifted in time by altering the step time. A total of twelve different input sequences were investigated, as shown in Fig. 5. For each sequence, ten different experiments have been performed and the average integral absolute errors (IAE) and integral square errors (ISE) of the model versus the plant were computed for the pitch and yaw velocities.

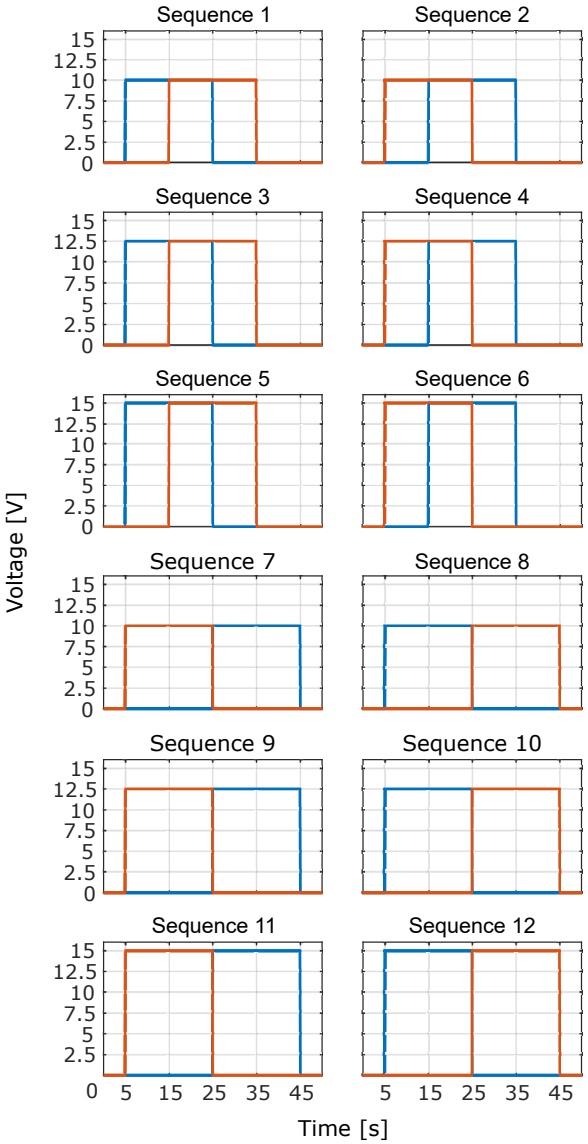


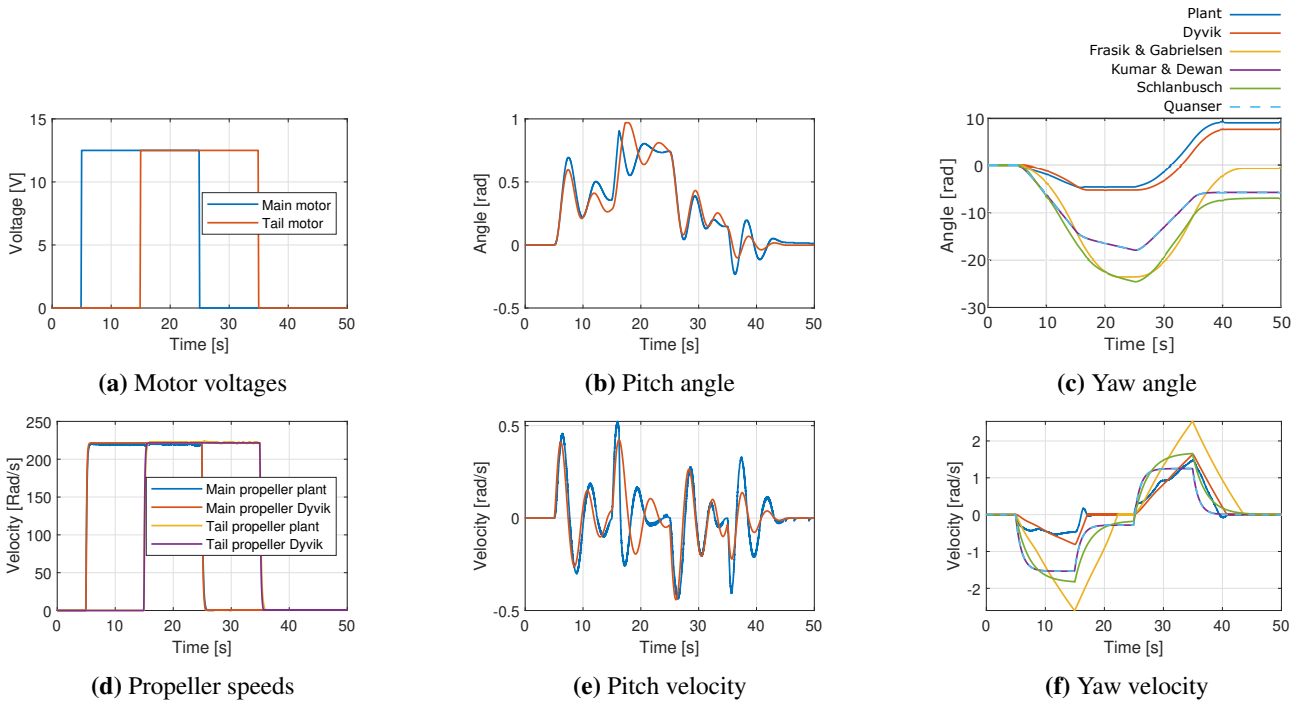
Figure 5. Input signal sequences

The performance of the model proposed in this paper compared to other state-of-the-art models in accurately predicting pitch angle, pitch velocity, yaw position, and yaw velocity, have been evaluated. Table 3 summarises the average IAE and ISE obtained for all input sequences. Fig. 6 shows the results in one particular run of input sequence number 3.

The results in the table show that, in the majority of experiments, the proposed model was either the best (dark green) or the second best (light green) model. Further examination of Table 3 reveals that the most significant improvement relates to the ability of the proposed model to predict yaw velocity. This is also evident in the plots presented in Figs. 6c and 6f. We can mention that the case of input sequence number 6, for which the yaw model underperformed, is the only observed exception. The improved performance of the proposed model with respect to predicting pitch motion, evident in Table 3, is primarily due to a more accurate representation of the frequency of oscillation and damping.

Table 3. Performance comparison of analysed models. The values are the average criteria after ten runs. Dark green color marks the highest performing model, and light green color marks the second highest performing model.

Sequence	Criteria		Frasik & Gabrielsen	Kumar & Dewan	Schlanbusch	Quanser	Dyvik & al.
1	dPitch	IAE	2.49	6.68	3.43	5.26	2.17
		ISE	0.28	1.47	0.39	1.07	0.21
	dYaw	IAE	14.95	20.84	24.63	20.84	3.94
		ISE	11.70	18.29	23.26	18.30	1.01
2	dPitch	IAE	3.07	6.64	3.44	4.94	2.19
		ISE	0.46	1.51	0.47	0.94	0.27
	dYaw	IAE	16.03	22.02	25.21	22.03	3.24
		ISE	12.78	20.30	24.49	20.31	0.66
3	dPitch	IAE	5.15	8.45	5.77	12.21	4.16
		ISE	0.94	2.38	1.55	5.88	0.77
	dYaw	IAE	30.62	20.61	21.80	20.62	6.73
		ISE	40.64	17.83	20.72	17.84	3.70
4	dPitch	IAE	6.62	8.52	4.28	8.65	3.83
		ISE	1.55	2.67	0.70	3.03	0.55
	dYaw	IAE	32.23	21.03	21.84	21.04	11.26
		ISE	42.76	19.33	22.28	19.34	6.53
5	dPitch	IAE	10.29	10.41	16.57	22.57	7.85
		ISE	4.04	3.98	12.19	25.36	2.70
	dYaw	IAE	48.51	26.81	21.75	26.82	8.13
		ISE	88.65	27.93	18.24	27.94	2.92
6	dPitch	IAE	10.86	9.22	16.23	19.24	6.23
		ISE	4.54	3.13	12.55	18.66	1.41
	dYaw	IAE	54.31	25.94	21.77	25.95	27.50
		ISE	101.07	27.04	18.39	27.06	41.20
7	dPitch	IAE	2.47	3.68	2.29	4.14	1.59
		ISE	0.31	0.58	0.25	0.75	0.16
	dYaw	IAE	46.80	39.83	42.84	39.84	11.15
		ISE	63.54	41.12	48.85	41.13	4.79
8	dPitch	IAE	2.22	3.44	1.97	3.45	1.76
		ISE	0.31	0.63	0.18	0.59	0.19
	dYaw	IAE	43.52	34.18	37.79	34.19	9.39
		ISE	56.18	33.95	43.36	33.96	2.59
9	dPitch	IAE	5.81	5.42	3.79	5.48	3.70
		ISE	1.43	1.28	0.68	1.51	0.77
	dYaw	IAE	71.47	39.67	40.42	39.68	23.00
		ISE	152.45	47.89	52.39	47.91	19.73
10	dPitch	IAE	4.77	4.61	3.75	5.64	3.01
		ISE	0.89	1.13	0.88	2.01	0.51
	dYaw	IAE	70.71	36.88	35.08	36.89	20.58
		ISE	170.88	37.67	40.93	37.68	13.51
11	dPitch	IAE	10.58	6.48	5.56	6.27	5.62
		ISE	3.86	2.01	1.44	2.48	1.41
	dYaw	IAE	99.85	51.09	47.52	51.10	42.14
		ISE	302.88	84.77	64.22	84.81	80.97
12	dPitch	IAE	10.31	6.09	4.65	5.06	3.05
		ISE	3.55	2.21	1.21	2.01	0.52
	dYaw	IAE	97.20	46.88	37.58	46.90	28.14
		ISE	320.08	71.30	41.64	71.33	25.43



(a) Motor voltages **(b)** Pitch angle **(c)** Yaw angle

(d) Propeller speeds **(e)** Pitch velocity **(f)** Yaw velocity

Figure 6. Experiment 3 results. Note that Figs. 6b and 6e contain only the proposed model due to the possible interpretational challenges that arise from the fact that all the considered models oscillate with slightly different frequencies and amplitudes.

5 Conclusions

This paper has proposed an improved model for describing the dynamical behavior of a Quanser Aero 2-DOF helicopter equipped with inefficient propellers. The model includes a more detailed description of friction and centripetal forces acting on the Aero. An extensive experimental validation using different input sequences has shown that the improved model provides better results, in the sense of lower values for IAE and ISE criteria, than other models proposed in the literature.

References

I. B. Abdelwahed, A. Mbarek, and K. Bouzrara. Model predictive control of 2-DOF helicopter based on MIMO ARX-Laguerre model. In *6th International Conference on Electrical and Electronics Engineering (ICEEE)*, pages 317–322, 2019.

R Christensen, N Fogh, RH Hansen, MS Jensen, S Larsen, and A Paramanathan. Modelling and control of a twin-rotor mimo system. *Department of Control Engineering Institute of Electronic Systems of Aalborg University*, 2006.

O. Egeland and J. T. Gravdahl. *Modeling and simulation for automatic control*. Marine Cybernetics Trondheim, 2002.

M. Fliess and C. Join. An alternative to proportional-integral and proportional-integral-derivative regulators: intelligent proportional-derivative regulators. *International Journal of Robust and Nonlinear Control*, 32(18):9512–9524, 2022.

J. M. Frasik and S. I. Lund Gabrielsen. Practical application of advanced control: An evaluation of control methods on a quanser aero. Master's thesis, Universitetet i Agder, 2018.

S.-K. Kim and C. K. Ahn. Performance-boosting attitude control for 2-DOF helicopter applications via surface stabilization approach. *IEEE Transactions on Industrial Electronics*, 69(7):7234–7243, 2021.

S. Kumar and L. Dewan. A comparative analysis of LQR and SMC for Quanser AERO. In *Control and Measurement Applications for Smart Grid: Select Proceedings of SGESC*, pages 453–463. 2022.

Quanser AERO - Laboratory Documentation. Quanser Inc, 2016.

H. Ren, Y. Zhao, W. Xiao, and Z. Hu. A review of UAV monitoring in mining areas: Current status and future perspectives. *International Journal of Coal Science & Technology*, 6:320–333, 2019.

S. M. Schlanbusch. Adaptive backstepping control of quanser 2DOF helicopter: Theory and experiments. Master's thesis, Universitetet i Agder, 2019.

E. Segerstrom, M. Podlaski, A. Khare, and L. Vanfretti. Parameter optimization and model validation of Quanser Aero using Modelica and RaPId. In *AIAA/IEEE Electric Aircraft Technologies Symposium (EATS)*, pages 1–9, 2021.

R. Shadiev and S. Yi. A systematic review of UAV applications to education. *Interactive Learning Environments*, pages 1–30, 2022.

C. A. Thiels, J. M. Aho, S. P. Zietlow, and D. H. Jenkins. Use of unmanned aerial vehicles for medical product transport. *Air medical journal*, 34(2):104–108, 2015.

Dynamic modelling and part-load behavior of a Brayton heat pump

Matteo Pettinari [a,*], Guido Francesco Frate [a], Konstantinos Kyprianidis [b], Lorenzo Ferrari [a]

[a] *University of Pisa, Department of Energy, Systems, Territory and Construction Engineering, Pisa, Italy*

[b] *Mälardalen University, Department of Sustainable Energy Systems, School of Business, Society and Engineering, Västerås, Sweden*

**corresponding author*: matteo.pettinari@phd.unipi.it

Abstract

Among the environmentally friendly technologies recently proposed in the literature, high-temperature heat pumps represent a promising solution to foster the complete penetration of renewables within the power grid. Such systems may be based on closed Brayton cycles and leverage many existing components. As they are meant to provide high-temperature heat while using renewable electricity, their potential field of application ranges from industrial heating to energy storage. Several variants are currently under development to assess the feasibility of such systems in providing flexibility to the electricity grid. To do so, they need to operate in part-load conditions and quickly react when the load must be adjusted. In this regard, this study investigates the transient capabilities of Brayton heat pump technology. To this extent, a detailed transient model of a novel prototype proposed in the literature is presented, accounting for controls, thermal inertia and volume dynamics related to heat exchangers and piping. Furthermore, the model is used to assess the transient performance of the system in response to sudden load variations, which is achieved by adapting the turbomachinery operating velocities. Results show that the system can safely operate in part-load conditions with regulation times compatible with industrial needs.

Keywords: high-temperature heat pump; dynamic modelling; transient simulation; control system.

Nomenclature

Abbreviations:

NTU	number of transfer unit, −
SM	surge margin, −
PR	pressure ratio, −
OV	overshoot, %
COP	coefficient of performance, −

Letter symbols:

t	time, s
$\dot{m}$	mass flow rate, kg/s
p	pressure, Pa
T	temperature, K
$\dot{T}$	temperature slope, K/min
ΔT	temperature difference, K
N	rotational speed, rpm
$\dot{Q}$	heat transfer rate, W
C	heat capacity rate, W/K

Greek symbols:

ε	effectiveness, −
τ	torque, N m

Subscripts and superscript:

0	reference state
1,2,3,..	cycle points
$corr$	corrected
r	ratio
min	minimum
max	maximum
hot	hot fluid/side
$cold$	cold fluid/side
$tube$	tube side
$shell$	shell side
$wall$	wall
$overall$	overall
$surge$	surge conditions
op	operating conditions
$rise$	rise
$settl$	settling

1. Introduction

According to the latest International Energy Agency report [1], heat production constitutes the world's largest energy end use, accounting for almost half of the global final energy consumption. Of the energy consumed for heat production in 2021, industry accounted for nearly 51 %, whilst 46 % was consumed for domestic purposes, such as

space and water heating in buildings, and 3 % in agriculture-related activities. Since fossil fuels still cover over 60 % of heating energy demand [1], heat production decarbonization represents a vital step towards carbon neutrality.

Among the solutions to decrease carbon dioxide emission due to heating, high-temperature heat pumps are increasingly recognized as a key technology [2]. Because of the progressive penetration of renewable energy sources, heat pumps may leverage excess renewable energy to recover heat from low-temperature sources (e.g., ambient heat, waste heat sources) and supply it to high-temperature heat utilizers (e.g. industrial processes or thermal energy storage), thus securing sustainable heating. Various heat pumps are available in the market, namely vapour compression, absorption, rotation, hybrid, and transcritical heat pumps [3]. The most common ones operate according to the Rankine thermodynamic cycle, although other cycles, such as Brayton heat pumps, were found to provide advantages depending on the process' source and sink temperature profile under consideration [4].

Arpagaus et al. [5] reported that most commercially available industrial heat pumps provide heat up to 90 °C, whilst only a few plants can supply heat at temperatures in the 140−160 °C range. Apart from domestic purposes, current fields of application comprise food, pulp and paper, metal, and chemical industries, as they benefit from low-temperature heat. Furthermore, several studies have shown a large demand for high-temperature heat in the European industry [6,7]. Therefore, such systems' range of utilization would likely extend in case higher heat temperatures were reached. For instance, heat pumps could serve for low-pressure steam production [5].

Low efficiencies, high equipment cost, long payback periods, component limitations, and the lack of environmentally friendly operating fluids performing adequately at high temperatures represent the main limitations to heat pump development [8]. However, Zühlsdorf et al. [9] suggest that heat pumps capable of supplying heat to at least 280 °C can be technically and economically feasible, for instance, by adopting equipment available in the oil and gas industry. Several concepts capable of delivering heat above 150 °C have been proposed in the literature for industrial purposes [5]. Further prototypes have also been proposed for energy storage applications [10]. Nevertheless, these heat pumps must be developed, built, and integrated into the relevant industries. Besides, such systems should also bring flexibility to the power grid, allowing for quick start-ups and load variations whilst operating safely and ideally at high efficiencies.

Despite being crucial for the demonstration of such technology, only a few works have so far focused on the transient capabilities of Brayton heat pumps. In particular, Frate et al. [11] investigated the power regulation of a Brayton PTES system through fluid inventory control, whilst other works analysed the behaviour of such systems during cold start-up [12,13]. As a contribution to this research topic, the paper investigates the transient capabilities of a novel Brayton heat pump. Particular focus is set on its response to sudden variations in the desired heat sink temperature by regulating the turbo-compressor operating speed.

2. Case study

The Brayton heat pump concept recently proposed by the German Aerospace Center (DLR) [12,14] and currently under realization in Cottbus, Germany, was considered in the present work. As a novel prototype, the heat pump aims to deliver heat at more than 250 °C. Figure 1 reports the plant layout, mainly comprising a three-stage turbo-compressor, three shell-and-tube heat exchangers, and a two-stage axial turbine.

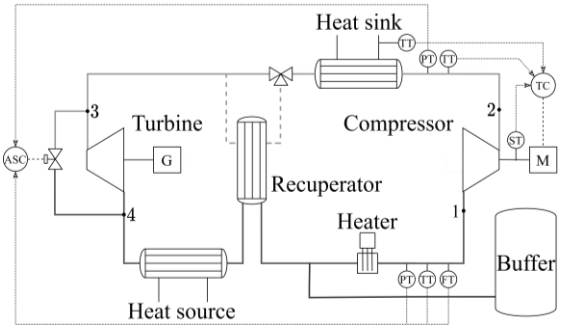


Figure 1: High-temperature heat pump layout.

The system operates according to the Brayton cycle. At first, the compressor increases a given operating fluid (e.g., dry air, Argon, or CO_2) temperature and pressure (pt. 1-2) whilst using electric power. The useful heat is then transferred to the thermal user employing the high-temperature heat exchanger (pt. 2-3), and afterwards, the fluid is expanded through the turbine to recover power and resultingly cool the fluid down to sub-ambient temperatures (pt. 3-4). A further heat exchanger transfers heat to the fluid from a low-temperature heat source (e.g., the environment) before returning it to the compressor (pt. 4-1). Moreover, a recuperator can recover heat internally, while a heater can also provide additional heat to raise the gas temperature at the compressor inlet. In the present work, the recuperator is assumed to be disabled, although present in the actual plant layout. In such a setup, the system supplies heat for 115 kW at 272 °C with a coefficient of performance (COP) of around 1.3. Finally, the system operates as a closed cycle, enabling fluid

inventory control to adjust the provided thermal load by extracting/injecting mass into the system from a secondary circuit (schematized as a buffer for simplicity). However, such a control strategy is not analysed in the present work.

3. Methodology

The heat pump was modelled within the Simulink® environment of Matlab® R2022b [15]. The Simscape™ package was used as it allows for easily implementing control systems due to the integration with Simulink®. For what concerns the operating fluids, dry air was considered in the closed cycle and secondary processes. It was modelled as a real gas, and its thermophysical properties were computed through RefProp [16].

3.1. Turbomachinery

The compressor and turbine were modelled at a system level and assumed adiabatic with negligible gas volume. The turbomachinery thermal dynamics were neglected, while rotor dynamics were considered by accounting for the compressor, turbine, motor, generator, and transmission moments of inertia.

As for the turbomachines' off-design modelling, actual performance maps provided by DLR were used, thereby specifying the relation between the conditions at the machine inlet (e.g., mass flow rate, pressure and temperature) with the component pressure ratio and the adiabatic efficiency. Figure 2 reports the compressor map used in the present work.

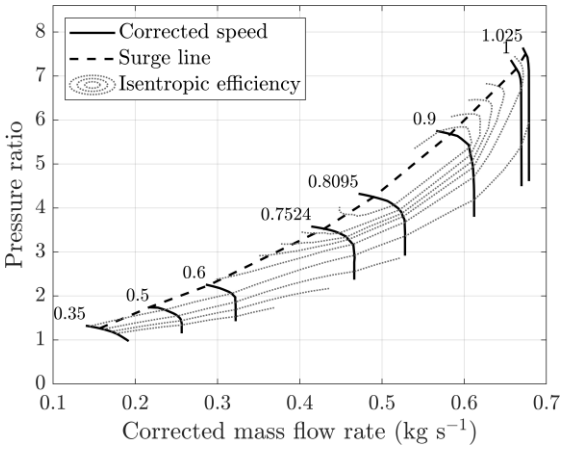


Figure 2: Compressor performance map [12].

It is worth noting that corrected parameters, defined as in Eq. (1), were used to account for pressure and temperature variations at the machine inlet:

$$\begin{cases} \dot{m}_{corr} = \dot{m}\dfrac{\sqrt{T/T_0}}{p/p_0} \\ N_{corr} = \dfrac{N}{\sqrt{T/T_0}} \end{cases} \quad (1)$$

where $p_0 = 101325$ Pa, $T_0 = 288.15$ K, N is the rotational speed, and $\dot{m}$, T, p are the inlet mass flow, temperature, and pressure, respectively.

3.2. Heat Exchangers

Similarly to the turbomachinery, the three shell-and-tube heat exchangers were assumed adiabatic and modelled based on the technical data provided by the manufacturer [12]. Therefore, characteristics such as geometry, flow arrangement, and metal mass were used to customize the components available in the Simscape™ Fluids library [17]. From a general standpoint, the flow conditions at each side of the heat exchangers were determined by means of mass, momentum, and energy balances. Conversely, the heat transfer rate exchanged between the two sides was established according to the ε-NTU methodology [18] as in Eq. (2):

$$\begin{cases} \dot{Q} = \varepsilon \cdot C_{min} \cdot \left(T_{hot,in} - T_{cold,in}\right) \\ NTU = \dfrac{1}{C_{min} R_{overall}} \\ \varepsilon = f(NTU, C_r) \end{cases} \quad (2)$$

where $\dot{Q}$ is the heat flow rate through the heat exchanger, ε is the heat exchanger effectiveness depending on the flow arrangement, C_{min} is the minimum heat capacity rate between the hot and cold side, C_r is the ratio between minimum and maximum heat capacity rate, and $R_{overall}$ is the overall thermal resistance of the heat exchangers.

Thermal inertia effects were further considered by modelling the thermal capacity due to the metal mass of the heat exchanger. The heat transfer rate at each side of the exchanger was then computed as:

$$\dot{Q}_{tube} = \dot{Q} + C_{wall,tube}\dot{T}_{wall,tube} \quad (3)$$

$$\dot{Q}_{shell} = \dot{Q} - C_{wall,shell}\dot{T}_{wall,shell} \quad (4)$$

where $C_{wall,i}$ is the thermal capacity of the heat exchangers evenly divided between the tube and shell sides, $\dot{T}_{wall,i}$ is the temperature slope in the wall half (positive if the temperature increases, negative when it drops), and $\dot{Q}$ is the heat transfer rate computed according to Eq. (2).

3.3. Piping, valves, and ancillary equipment

Piping's contribution to the system gas volume and its thermal inertia was accounted for in the model. In particular, pipes were assumed adiabatic, and their volumes and metal mass were determined based on their actual geometry. Similarly, the turbine bypass and recuperation valves were modelled as adiabatic local restrictions with negligible volume. Their geometry, as well as their flow characteristic, were specified based on

manufacturer data. Finally, only pressure drops were considered for what concerns the heater.

3.4. Control strategy

Two control loops were designed to safely operate the system in transient conditions, namely:

- anti-surge controller;
- temperature controller.

The former prevented the compressor from operating within critical areas near the surge line. In more detail, the compressor surge was avoided by varying the turbine bypass opening to increase the compressor flow, as shown in Figure 1. Moreover, the valve opening was determined by a variable-gain PI feedback controller based on the surge margin residual. In the present work, a minimum surge margin of 10 % was ensured through the anti-surge regulator and computed as:

$$SM = \left(\frac{PR_{surge}}{PR_{op}} - 1 \right)_{\dot{m}_{corr}} \quad (5)$$

where PR_{op} is the operating pressure ratio, whilst PR_{surge} is the surge line pressure ratio evaluated at the same corrected mass flow.

On the other hand, the temperature controller was used to regulate the sink outlet temperature at the desired setpoint by properly varying the motor/compressor speed. As the purpose of the analysis was to investigate the system response and not to optimize the system performance, only the following control requirements were considered:

- zero steady-state residual due to variations in the sink outlet temperature setpoint;
- temperature slopes of no more than 2 K/min to avoid critical thermal stresses at the heat exchangers [12].

A cascade architecture was considered to meet the considered requirements. Figure 3 reports the control schematic.

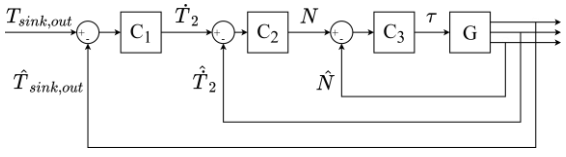


Figure 3: Temperature control schematic.

Three control loops were implemented, each one characterized by a time-continuous PI regulator. The outer loop was used to determine a setpoint for the compressor outlet temperature slope based on the sink outlet temperature residual. Here, the slope setpoint was limited to a maximum value of ± 2 K/min to avoid undesired stresses at the high-temperature heat exchanger. The intermediate control loop then compares the temperature slope setpoint with its measurement (or estimate), thus providing a motor speed setpoint. In analogy to the outer loop, the speed setpoint was limited to guarantee the compressor rotates at admissible shaft speeds. Lastly, the inner controller determined the motor torque based on the speed error. In this regard, the motor speed ramp limits were not accounted for within the control architecture for simplicity. However, the resulting shaft accelerations were verified a posteriori to ensure they did not exceed the maximum limit of 300 rpm/s. Finally, anti-windup methods were adopted to prevent integration wind-up of the regulators.

4. Results and Discussion

4.1. Performed analyses

The heat pump model presented in the previous sections was used to study the system behaviour in response to sudden variations of the sink outlet temperature demand.

A sink outlet temperature increment of 10 °C was considered at first to investigate the system's response in detail. Such a simulation aims at assessing the characteristic response time of the system and the main issues that occur when the high-temperature secondary process requires a given mass flow but at a higher temperature with respect to the nominal one. Secondly, a sensitivity analysis was performed by varying the sink outlet temperature setpoint of up to 20 °C. Here, the primary purpose was to explore the operability range of the system when driven by such a regulation approach. Finally, a further analysis was performed to assess the impact of the maximum allowed thermal stresses at the heat exchangers on the system response by simulating the system for the same required sink outlet temperature lift and different discharge temperature slope limits of the compressor.

At the beginning of each scenario, the system was assumed to operate at nominal, steady conditions, delivering heat for 115 kW at 272 °C to the secondary process, raising its fluid temperature to 261.4 °C. Therefore, all the components were assumed already in steady-state temperature, and turbomachines rotating at their nominal speeds. Finally, it is worth mentioning that control actuators such as the heater and the recuperation three-way valve were not utilised in the present analyses, despite being accounted for in the model. In particular, the three-way valve was regulated such that the primary mass flow exiting the high-temperature heat exchanger went directly to the turbine inlet.

4.2. Setpoint variation

The system response due to a step change in the sink outlet temperature demand is shown in Figure 4. At the beginning of the simulation, the

system operates at nominal conditions, heating the sink flow from 15 °C to around 261 °C. At t = 5 min, the desired temperature at the sink outlet is suddenly increased by 10 °C. As reported in Figure 4b, to accommodate the setpoint movement, the temperature controller accelerates the compressor, which provides a higher and hotter mass flow rate at the high-temperature heat exchanger. As the primary flow and its temperature increase, more heat is transferred to the secondary process, and the sink mass flow outer temperature rises to the desired value.

From a transient perspective, the sink outlet temperature evolves as a second-order response characterized by a settling time of 12.5 min and an overshoot of 9.5 %. Moreover, during the rise time (t_{rise} = 6.45 min), it can be observed that the sink outlet temperature increases almost constantly due to the temperature controller. As reported in Figure 4b, to prevent thermal stresses at the heat exchanger, the controller accelerates the compressor while ensuring its outer temperature (e.g. T_2) does not rise with slopes higher than 2 K/min. In particular, both the relative speed and the sink outlet temperature vary linearly when the maximum allowed slope is reached. As the temperature slope limit bounds the machine acceleration, it can be stated that it represents a critical parameter characterizing the system response to sudden temperature variations. Allowing for higher slopes by adopting more advanced heat exchangers may likely shorten the rise time.

On the other hand, nearly before t = 15 min, the measured sink outlet temperature experiences some oscillations, which extend the response time to around 3−4 min. Here, both overshoots and undershoots are primarily due to the thermal inertia of the heat exchangers and piping. When optimizing the system response, adopting more sophisticated regulators with a feedforward component (e.g., FF-FB or MPC) may be considered to dampen the oscillations, thus further shortening the response time. From t = 20 min until the end of the simulation, the compressor speed is slowly adjusted as the temperatures within the heat exchangers and piping settle.

It is also interesting to observe how the system performance varies during the simulation. Figure 4c reports the plant COP and the main quantities involved in its calculation. It can be observed that the COP decreases from around 1.30 to 1.18 during the regulation and finally settles to 1.27 as the new operating conditions are met. This is strictly related to the control strategy under analysis that varies the compressor operating point. Since the compressor rotates at higher speeds, it operates at lower efficiencies, absorbing more power. Although the exchanged thermal power and

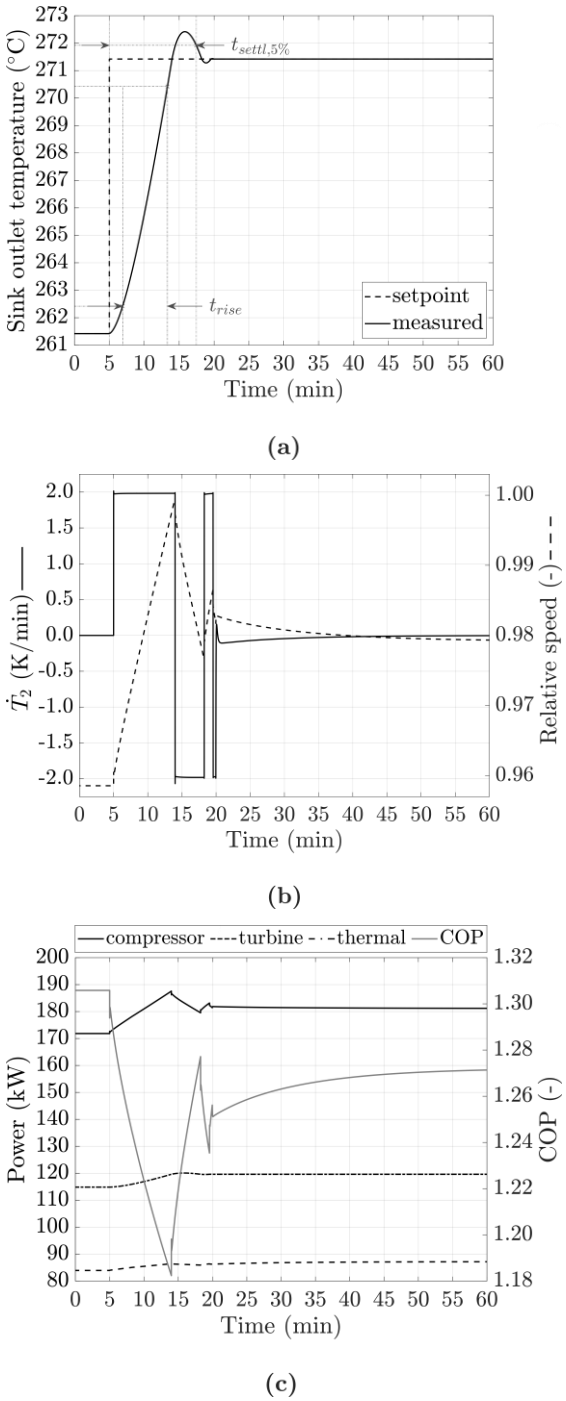


Figure 4: System response to a variation of the desired sink outlet temperature: (a) sink outlet temperature; (b) control signals; (c) system performance.

the power recovered by the turbine also increase due to the higher and hotter mass flow rate circulating within the closed cycle, they vary less significantly than in the compressor. Therefore, the overall system COP is lower. Finally, it is worth noting the differences among the compressor, turbine, and thermal power exchanged through the high-pressure heat exchanger during the regulation of the system (t = 5–20 min). In particular, the heat exchanger's thermal inertia helps mitigate the

abrupt variations of the power absorbed by the compressor, resulting in smoother profiles for both the exchanged thermal power and the power recovered by the turbine.

4.3. Sensitivity to the desired sink temperature lift

A sensitivity analysis was performed to compare the system behaviour for different temperature setpoint variations and to investigate the adopted control strategy's limit. Results are shown in Figure 5, which reports the sink outlet temperature, the measured temperature slope at the compressor outlet and the motor/compressor relative speed. Table 1 summarizes the main properties characterizing the sink outlet temperature responses instead.

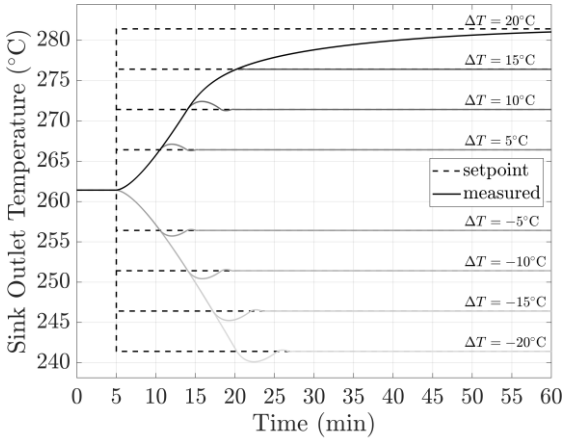
Table 1: Sink outlet temperature response characteristics for different desired temperature steps.

ΔT (°C)	t_{rise} (min)	$t_{settl,5\%}$ (min)	OV (%)
-20	10.95	18.73	6.47
-15	8.75	15.82	7.82
-10	6.47	12.52	9.88
-5	3.93	8.5	14
5	3.93	8.5	13.92
10	6.45	12.5	9.97
15	9.91	13.53	0.41
20	25.8	40.85	-

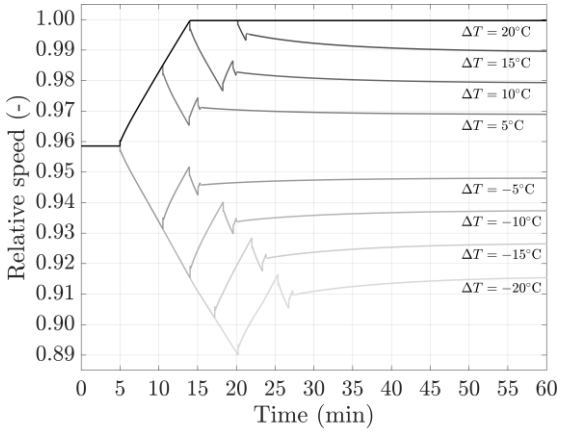
In each considered scenario, it can be observed that both the compressor speed and the sink outlet temperature constantly increase as the setpoint perturbations occur due to temperature control. As discussed in the previous section, the rate at which these quantities rise depends on the maximum thermal stress sustainable by the heat exchangers. Furthermore, except for temperature increments of 15−20 °C, each response oscillates and stabilizes at the desired value.

It can be noted that the system response is symmetrical with respect to the nominal conditions. As reported in Table 1, in the case of changes of equal magnitude, the system exhibits almost identical response times to reach the new operating conditions. The oscillation magnitude is also similar, although it is the sign of the variation to determine whether the sink outlet temperature undergoes an overshoot or an undershoot. Different behaviour is observed in the case of increments higher than ten degrees. For a required temperature step of 15 °C, the sink outlet temperature does not experience any overshoot and shows a longer rise time and shorter settling time compared to a temperature variation of −15 °C. On the other hand, the system takes more than 40 min to increase the sink outlet temperature of 20 °C, and zero steady error is reached slightly after one hour.

These differences are related to how faster the compressor can operate compared to the nominal



(a)



(b)

Figure 5: Sensitivity to the temperature setpoint variation: (a) measured sink outlet temperature; (b) relative motor/compressor speed.

operating point, which primarily depends on its performance map. As reported in Figure 5b, in such cases, the relative speed of the motor/compressor continuously increases until it reaches the maximum limit of 1, meaning that the compressor is operating at its maximum speed and cannot increase its outlet temperature any further. Consequently, since the compressor cannot provide higher and hotter mass flow at the heat exchanger, more time is required for the heat exchanger material to warm up and overcome its thermal inertia, eventually leading to a rise in the sink outlet temperature. Such behaviour also implies that higher sink outlet temperature step increments are not achievable by the system through the proposed regulation strategy.

Conversely, such issues do not occur in the case of a desired temperature reduction at the sink outlet as the compressor decelerates to reduce the heat transferred through the high-temperature heat exchanger, thus decreasing the sink outlet temperature. Here, depending on the temperature step magnitude (e.g., ΔT = −15,−20 °C), it is worth

pointing out that the anti-surge regulator may be enabled to avoid compressor surge, thus ensuring the system operates safely.

4.4. Sensitivity to the maximum temperature slope

The system response to a desired sink outlet temperature step of 10 °C is reported in Figure 6. Different maximum compressor outlet temperature slopes are considered, which results in higher thermal stresses at the heat exchangers being allowed.

As shown in Figure 6a and Table 2, higher slopes lead to quicker response times. In particular, both the rise and settling times reduce significantly when a temperature slope of 5 K/min is permitted. Further improvements are achieved for slopes of 10−15 K/min, although less relevant (e.g., the settling time difference is around 1 min when comparing results obtained using slopes of 5 K/min and 10 K/min, whilst less than 20 s for slopes of 10 K/min and 15 K/min). The transients' quality also improves when higher thermal stresses are allowed, as the response maximum overshoot reduces by one-half for 5 K/min slopes and becomes negligible for 10−15 K/min.

Table 2: Sink outlet temperature response characteristics for different allowed temperature slopes at the compressor outlet.

$\dot{T}_{2,max}$ (K/min)	t_{rise} (min)	$t_{settl,5\%}$ (min)	OV (%)
2	6.45	12.5	9.97
5	3.68	5.06	4.01
10	3.08	4.13	1.25
15	2.98	3.82	0.51

From a general standpoint, the higher slopes lead to quicker responses as they permit steeper speed ramps, as reported in Figure 6b. As a result, the compressor can provide higher mass flow at the maximum reachable temperature sooner at the high-temperature heat exchanger, thus shortening the time required to raise the sink outlet temperature. The same also holds for the response oscillations, which become lower and lower as higher temperature slopes at the compressor outlet are considered. However, despite allowing for quicker responses, almost negligible improvements are observed when increasing the maximum slope from 10 K/min to 15 K/min, which suggests that using heat exchangers capable of sustaining higher thermal gradients may not be practical, as they would likely be more complex and expensive whilst providing similar performance.

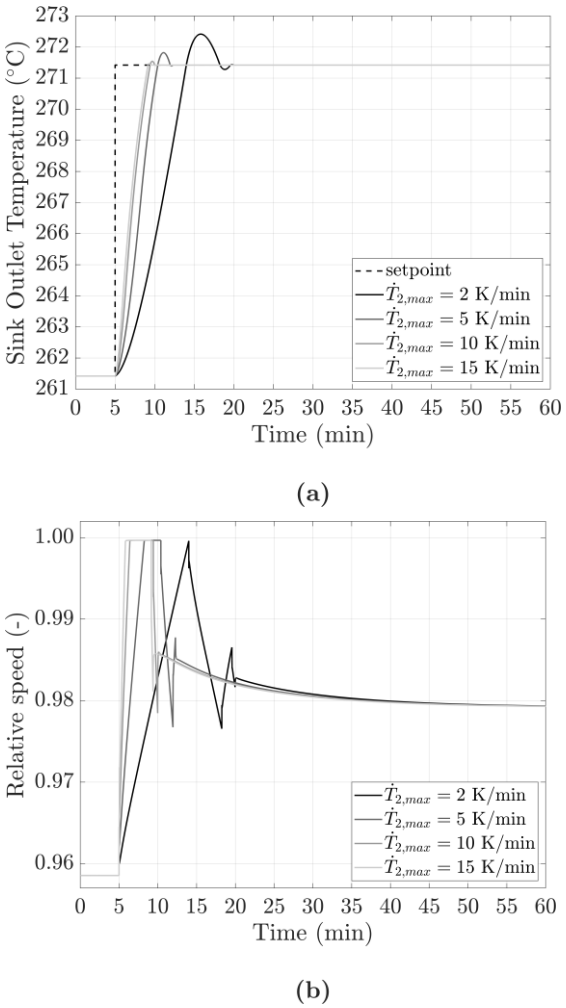


Figure 6: Sensitivity to the compressor outlet maximum allowed temperature slope: (a) measured sink outlet temperature; (b) relative motor/compressor speed.

5. Conclusions

The paper analyzed the transient behaviour of a Brayton heat pump responding to a desired temperature variation at the sink outlet by varying the compressor operating speed. To this extent, a detailed system model was developed, and control loops were designed to safely operate the system in transient conditions. The model was then used to investigate the system response for a required sink outlet temperature increment and to analyze the range of applications of the considered control strategy. Results can be summarized in the following statements:

- the system raises the sink outlet temperature to the desired value with a second-order system response. The response is symmetrical with respect to the nominal operating conditions;
- the rate at which the sink outlet temperature increases is limited by the maximum allowed temperature slope at the compressor outlet. Heat exchangers capable of sustaining higher thermal stresses are then required to reduce the system response time;
- oscillations characterizing the system response are primarily due to the system's thermal inertia. More sophisticated controllers may be

considered to avoid overshoots, further improving the system transient;
- the maximum sink outlet temperature achievable by the heat pump is limited by the maximum operating speed of the compressor, which also affects the system response time for desired temperature lifts higher than 10 °C.
- allowing for higher thermal stresses at the heat exchangers leads to faster system responses, although the improvements progressively decrease for temperature slopes higher than 5 K/min.

Acknowledgment

The authors would like to acknowledge Johannes Oehler, Anh Phong Tran, and Panagiotis Stathopoulos from DLR for providing detailed data concerning the Brayton heat pump under analysis and for their support and feedback.

This research has received financial contribution from the Italian Operative National Plan (Piano Operativo Nazionale, PON) in the framework of the project Ricerca e Innovazione 2014−2020 (PON R&I) – Azioni IV.4 e IV.5 “Dottorati di ricerca su tematiche dell'innovazione e green” (DM MUR 1061/2022) e IV.6 “Contratti di ricerca su tematiche dell'innovazione e green” (DM MUR 1062/2022).

References

[1] IEA, 2022, *Heating*, IEA, Paris, https://www.iea.org/reports/heating (last accessed: 30/05/2023)

[2] IEA, 2022, *The Future of Heat Pumps*, IEA, Paris, https://www.iea.org/reports/the-future-of-heat-pumps (last accessed: 30/05/2023).

[3] Wolf, S., and Blesl, M., 2016, “Model-Based Quantification of the Contribution of Industrial Heat Pumps to the European Climate Change Mitigation Strategy,” *ECEEE Industrial Summer Study Proceedings*, Berlin, Germany, pp. 477–487.

[4] Gai, L., Varbanov, P. S., Walmsley, T. G., and Klemeš, J. J., 2020, “Critical Analysis of Process Integration Options for Joule-Cycle and Conventional Heat Pumps,” Energies, **13**(3), p. 635. https://doi.org/10.3390/en13030635.

[5] Arpagaus, C., Bless, F., Uhlmann, M., Schiffmann, J., and Bertsch, S. S., 2018, “High Temperature Heat Pumps: Market Overview, State of the Art, Research Status, Refrigerants, and Application Potentials,” Energy, **152**, pp. 985–1010. https://doi.org/10.1016/j.energy.2018.03.166.

[6] Naegler, T., Simon, S., Klein, M., and Gils, H. C., 2015, “Quantification of the European Industrial Heat Demand by Branch and Temperature Level: Quantification of European Industrial Heat Demand,” Int. J. Energy Res., **39**(15), pp. 2019–2030. https://doi.org/10.1002/er.3436.

[7] Rehfeldt, M., Fleiter, T., and Toro, F., 2018, “A Bottom-up Estimation of the Heating and Cooling Demand in European Industry,” Energy Effic., **11**(5), pp. 1057–1082. https://doi.org/10.1007/s12053-017-9571-y.

[8] Jesper, M., Schlosser, F., Pag, F., Walmsley, T. G., Schmitt, B., and Vajen, K., 2021, “Large-Scale Heat Pumps: Uptake and Performance Modelling of Market-Available Devices,” Renew. Sustain. Energy Rev., **137**, p. 110646. https://doi.org/10.1016/j.rser.2020.110646.

[9] Zühlsdorf, B., Bühler, F., Bantle, M., and Elmegaard, B., 2019, “Analysis of Technologies and Potentials for Heat Pump-Based Process Heat Supply above 150 °C,” Energy Convers. Manag. X, **2**, p. 100011. https://doi.org/10.1016/j.ecmx.2019.100011.

[10] Smith, N. R., Tom, B., Rimpel, A., Just, J., Marshall, M., Khawly, G., Revak, T., and Hoopes, K., 2022, “The Design of a Small-Scale Pumped Heat Energy Storage System for the Demonstration of Controls and Operability,” *Proceedings of the ASME Turbo Expo 2022: Turbomachinery Technical Conference and Exposition*. *Volume 4: Cycle Innovations; Cycle Innovations: Energy Storage*. Rotterdam, Netherlands. V004T07A012. ASME. https://doi.org/10.1115/GT2022-83424

[11] Frate, G. F., Pettinari, M., Di Pino Incognito, E., Costanzi, R., and Ferrari, L., 2022, “Dynamic Modelling of a Brayton PTES System,” *Proceedings of the ASME Turbo Expo 2022: Turbomachinery Technical Conference and Exposition*. *Volume 4: Cycle Innovations; Cycle Innovations: Energy Storage*. Rotterdam, Netherlands. V004T07A013. ASME. https://doi.org/10.1115/GT2022-83445.

[12] Oehler, J., Tran, A. P., and Stathopoulos, P., 2022, “Simulation of a Safe Start-Up Maneuver for a Brayton Heat Pump,” *Proceedings of the ASME Turbo Expo 2022: Turbomachinery Technical Conference and Exposition*. *Volume 4: Cycle Innovations; Cycle Innovations: Energy Storage*. Rotterdam, Netherlands. V004T06A003. ASME. https://doi.org/10.1115/GT2022-79399.

[13] Pettinari, M., Frate, G. F., Tran, A. P., Oehler, J., Stathopoulos, P., and Ferrari, L., 2023, “Transient Analysis and Control of a Brayton Heat Pump during Start-Up,” *Proceedings of ECOS 2023: The 36th International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems*, Las Palmas de Gran Canaria, Spain, pp. 839–850. https://doi.org/10.52202/069564-0076.

[14] Oehler, J., Gollasch, J., Tran, A. P., and Nicke, E., 2021, “Part Load Capability of a High Temperature Heat Pump with Reversed Brayton Cycle,” *13th IEA Heat Pump Conference*, Jeju, Korea, p. 12.

[15] The MathWorks Inc., 2022, “MATLAB Version: 9.13.0 (R2022b),” https://www.mathworks.com (last accessed: 30/05/2023)

[16] Lemmon, E. W., Bell, I. H., Huber, M. L., and McLinden, M. O., 2018, “NIST Standard Reference Database 23: Reference Fluid Thermodynamic and Transport Properties-REFPROP, Version 10.0, National Institute of Standards and Technology.” https://doi.org/10.18434/T4/1502528.

[17] The MathWorks Inc., “Simscape Fluids Reference,”https://it.mathworks.com/help/pdf_doc/hydro/hydro_ref.pdf (last accessed: 30/05/2023).

[18] Holman, J. P., 2002, *Heat Transfer*, 9th ed., McGraw-Hill, New York, NY, ISBN: 978-0-07-352936-3.

Future Potential Impact of Wind Energy in Sweden's bidding area SE3

J.J.Warners[a] S. Vouros[b,*] K. Kyprianidis[b] R. Benders[a] P. Nienhuis[a]

[a]*University of Groningen,* [b]*Mälardalen University*
[*]stavros.vouros@mdu.se

Abstract

This research addresses the potential for increasing wind power in Sweden's bidding area SE3. Sweden currently faces an energy imbalance, with larger production in the north and high demand in the south. Four bidding areas were introduced to incentivize energy production in the south. SE3, the largest bidding area, represents 60% of total demand. Using Seasonal Auto-Regressive Integrated Moving Average (SARIMA), historic data analysis from 2007 to 2022 is forecasted to a medium long-term future of 2035. Forecasting the observed trends reveals a potential supply deficit even under minimum demand growth scenarios made in literature. Closure of nuclear plants contributes to the shortfall, and the increasing trend in solar and wind power falls short. To study the impact wind power can have, the monthly wind patterns are analyzed, and used to calculate the power potential of different turbine capacities. Offshore areas show the highest potential for increasing wind power capacity in SE3. Economic factors, like payback time, are considered. The research concludes that there is technically and economically viable potential for wind power capacity to address the demand-supply gap by 2035. However, it depends on permitted areas, excluding built areas, UNESCO sites, and fishing routes. Future research should further explore these restrictions and address the seasonal variability in wind power to improve the understanding of the potential for wind power in the SE3 bidding area.

1 Introduction

Swedish electricity generation has historically been reliant on nuclear and hydro power and produces very little emissions. However, Swedish nuclear power is being phased out, and it is uncertain how many reactors will be operational after 2040 (SWEA, 2021; IEA, 2019). An expansion of wind power can replace fossil fuels and nuclear power and thus contributes positively to the environmental quality goals (Energimyndigheten, 2018; SWEA, 2021). These goals follow from the Energy Agreement formed by the parliament in 2016 and got adopted in 2018 (IEA, 2019). It states that the energy policy in Sweden should combine ecological sustainability, competitiveness and security of supply (IEA, 2019). Regarding electricity, Sweden must have a network with high security of supply and low environmental impact and offer electricity at competitive prices (IEA, 2019; Statens Energimyndighet, 2022). Relative environmental targets that result from the agreement are (IEA, 2019; Råberger & Vingmarker, 2019; Statens Energimyndighet, 2022): net zero emissions by 2045 and negative thereafter, 63% lower greenhouse gas emissions by 2030 compared to 1990, energy use must be 50% more efficient in 2030 compared to 2005, and a 100% renewable electricity production by 2040 (excluding nuclear energy). The Energy Agreement states that Sweden must have a network with high security of supply and low environmental impact and offer electricity at competitive prices (IEA, 2019; Statens Energimyndighet, 2022).Increasing the amount of wind power contributes to creating a more resilient and secure energy network because of the spatial distribution of power plants and would also meet the urgent power demand in southern Sweden (SWEA, 2022). However, at the same time, the increased share of renewables in the power grid raises concerns about security due to the intermittent characteristics in the availability of wind and solar energy (Gawel et al., 2017). Sweden is subdivided into four bidding areas, SE1-4 (Svenska Kraftnät, 2022). Even though Sweden is producing most of its energy within its borders, southern areas SE3 and SE4 have a relative energy deficit compared to the northern areas SE1 and SE2 (Armelius, 2022; IEA, 2019). The difference in price between the bidding areas was introduced in 2011 to encourage an increase in power production capacity in the southern areas (IEA, 2019). SE3 is the largest energy consumer, as it contains the largest share of the population and industry with cities such as Stockholm and Göteborg (Svenska Kraftnät, 2007-2022; Svenska Kraftnät, 2021; Back, 2020). The Swedish wind energy agency (SWEA) estimates an increase from 30TWh in 2020 to 60TWh with a capacity of 18.5GW in 2030 and 120TWh with a capacity of 33.3GW in 2040 (SWEA, 2021). The overall objective of this report is to analyze historical electricity consumption and production data, study wind patterns, determine payback times for turbines, and assess the po-

tential for wind power to meet future demand. The goal is to create a tool for evaluating feasible wind farm locations and assessing the impact on the supply-demand balance in the SE3 area. The research aims to contribute to higher electrical independence, lower consumer prices, and a better understanding of wind power's role in the energy mix (SWEA, 2022). This is done with the aim of answering the question: *How complementary can wind power be in the SE3 bidding area?*

2 Methodology

To be able to understand the impact wind power can have in the SE3 area, multiple parameters need to be studied. Historic energy data in SE3 needs to be understood before setting future demand scenarios. This data is obtained from the Transmission System Operator (TSO) Svenska Kraftnät and completed using the imbalance settlement services from eSett. Analysis of past trends, combined with literature on supply and demand changes, enables estimations of future supply-demand gaps. Forecasting models, such as the seasonal auto-regressive integrated moving average (SARIMA), will be employed to establish scenarios based on historic patterns. Assuming a linear increase in electricity demand and a stable share of Sweden's total consumption in SE3 of 60%, minimum and maximum demand growth scenarios will be considered. Next to this, wind data analysis using weather station data around SE3 will assess the potential of wind power to address power production deficiencies and achieve climate goals. The study will calculate turbine payback time based on power generation capacity, considering both onshore and offshore locations. Gaussian process regression will be used to fill gaps between weather stations and create a comprehensive wind speed map for SE3. Surface roughness lengths extracted from the Global Wind Atlas will be used to convert wind speed measurements from weather stations to turbine height (Badger et al., 2023). The grid density obtained from the surface roughness length data results in cells being approximately 1 by 1.5 kilometres. Considering current and future turbine capacities, the research assumes an onshore capacity increase from 3.5 MW to 15 MW and offshore capacity increase from 12.5 MW to 25 MW for calculations in 2035 (SWEA, 2021). Evaluating power calculations for different turbine capacities will aid in identifying suitable locations. The impact on the grid will be assessed based on the payback time of each turbine, with cost being a determining factor for installation and grid integration. To assess the payback time, the historic day-ahead electricity prices will be observed, which is obtained from European Network of Transmission System Operators for Electricity (ENTSO-E) (ENTSO-E, 2023). The electricity price observed will be used to set a value for the electricity price when calculating the payback times. This will not be forecasted due to the large variability and dependencies in the electricity price determination.

2.1 Wind power calculations

To be able to comprehend the wind speeds attributes, understanding the wind speed dependency of turbines is crucial. The theoretical energy that can be harvested from wind can be determined using (Boyle, 1996; Shu & Jesson, 2021):

$$P = \frac{1}{2} A \rho u^3 \tag{1}$$

Where P is the energy, A the swept area of the rotor, ρ the air density and u the wind speed. The unpredictability of wind characteristics has resulted in a probability distribution that describes the variation of wind speed at a location, indicating the likelihood of a wind speed to occur at a location (Shu & Jesson, 2021; Burton et al., 2011). The two-parameter Weibull distribution is most commonly used to perform the wind energy assessment and capture the skewness of the wind speed distribution (Shu & Jesson, 2021; Burton et al., 2011). The Weibull distribution is a function of wind speed $f(u)$;

$$f(u) = \left(\frac{k}{c}\right)\left(\frac{u}{c}\right)^{k-1} \exp\left[-\left(\frac{u}{c}\right)^{k}\right] \quad (u > 0; k, c > 0) \tag{2}$$

The function is further described by a scale parameter (c (m/s)), which establishes the horizontal axis scale of the wind distribution and a dimensionless shape parameter (k), describing the width of the distribution. c is thus linked to the wind speeds at a site, or elevation. Meaning, determining the ratio in speed at different altitudes can help to determine the c parameter at another altitude once it is calculated at an elevation. This leads to a shift in the probability density function described by the Weibull distribution, but the shape of the function, described by k would not change. The probability of a turbine being able to operate in the area can be determined using (2), and can be written as (Shu & Jesson, 2021):

$$p(u) = \exp\left[-\left(\frac{u_c}{c}\right)^{k}\right] - \exp\left[-\left(\frac{u_f}{c}\right)^{k}\right] \quad (u_c < u < u_f) \tag{3}$$

With the cut-in wind speed (u_c), the cutoff wind speed (u_f). The total energy a turbine produces over a certain time can be calculated by:

$$E = \sum_{0}^{i} (P_i \cdot p_i \cdot dv_i \cdot t) \tag{4}$$

where P_i is the power at a wind speed, p_i is the probability of that wind speed occurring according to the Weibull distribution, dv_i is the interval of the wind speeds in the power curve from the turbine and t is the time over which the turbine runs. Which in turn can be used to determine the capacity factor (CF) of the turbine. This describes the ratio of measured energy a turbine produced to its rated energy over a year.

2.2 Research boundaries

The Swedish electricity grid faces barriers in electrification related to generation, system adequacy, demand, and grid infrastructure (DNV, 2021). In the SE3 area, the lack of local generation and available grid pose significant challenges, along with lengthy planning and permitting processes for new projects and grid upgrades (Armelius, 2022; DNV, 2021). The intermittent behaviour of renewable sources like wind and solar adds pressure to system adequacy, highlighting the need for storage and grid flexibility (DNV, 2021). However, this research does not specifically address storage implementation or the use of battery electric vehicles for grid balancing. Factors such as population growth, economic development, technological advancements, and political decisions impact electricity demand but are not individually addressed in this study. Demand forecasts from the literature, representing minimum and maximum growth scenarios, will be considered in conjunction with supply forecasts using SARIMA. The research focuses on providing technologically feasible results that account for seasonal changes and contribute to addressing the power production gap with wind power. The study does not consider the influence of political decisions, public opinions, or other external factors. Limitations on possible wind turbine sites such as jurisdictional and natural factors, as well as stakeholder influence are not fully examined. Economic indicators, specifically the payback time, are considered in determining turbine feasibility, while grid parameters are not included. Increasing north-to-south transmission capacity to utilize wind energy in northern areas is an alternative to reduce SE3's self-supply requirements.

3 Results and Discussion

Following the methods on the data gathered different results are obtained. The historical data is transformed and visualized after which the production sources are forecasted. Data from weather stations are transformed to show the general wind direction over the region and the average monthly wind speeds. This in turn can be used to calculate the power a turbine can produce in the area, after which the payback time for a turbine can be calculated. The results obtained from the forecasting can be compared with the technical potentially available wind power with a reasonable payback time to learn about the role wind power can play in SE3.

3.1 Electricity demand and supply

The data from 2007 to 2022 is used to observe the historic behaviour in demand and supply. The behaviour in the different supply sources is forecasted using SARIMA in order to keep the trend. The trend observed in demand will be compared with different demand scenarios in 2035 obtained from literature. Combining these forecasted results gives an insight in the future demand and supply of the SE3 area for the different scenarios and when following the current trends in power supply.

3.1.1 Historic demand and supply

The annual consumption from 2007 to 2022 is visualized in Fig. 1. The average annual consumption during this time period is found to be 83.3 TWh. The stability in consumption can be explained by trends counteracting each other; in the 1980s and 1990s the demand increased by a growth in electric heating, but the shift to efficient heat pumps stopped the growth in electricity demand, and also the electrification of the industry has been counteracted by increasing efficiencies (IEA, 2019).

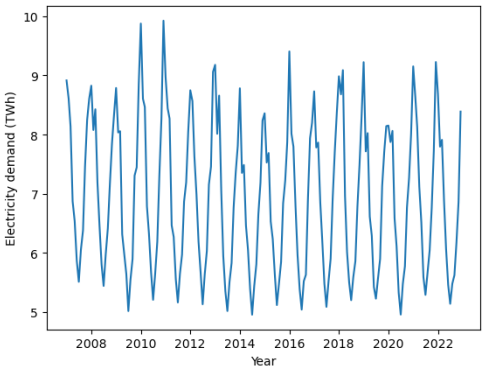


Figure 1. The total monthly consumption in SE3 from 2007 to 2022.

Fig. 2 shows how the demand is met by the different types of power production sources.

Nuclear power is seen to form the large base in the region and gets supplemented by hydro power to make up the vast majority of the power production. Thermal power can be seen to mostly add in production during the winter months. The increase in wind power can be observed, whereas the increase in solar power is still to small too be visible. The reduction in production from nuclear power is the result of the decommissioning of nuclear power plants in 2020. The forecasting

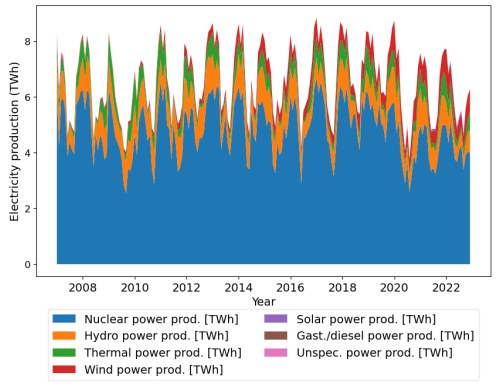


Figure 2. Total monthly electricity production by source in SE3 from 2007 to 2022.

of nuclear power assumes a continuation in production based on the capacity from 2020 onwards. Hydro and thermal power have been a stable source throughout these years and are assumed to stay stable when forecasting these sources. The observed solar power production shows an exponential growth up to 2022. The forecast follows a linear increase, which is in line with the expectation of the total installed power from 2023 to 2030 for the whole of Sweden (JB Sustainable Approach AB, 2019). This linear growth is assumed to continue to 2035. The wind power production has a linear increase, which is continued.

3.1.2 Forecasting demand and supply

The supply is forecasted using SARIMA based on the trends observed and explained in Section 3.1.1. To be able to adjust the demand growth to the different scenarios, the demand from Fig. 1 is represented by a sine wave. The seasonal variability in consumption is assumed to stay the same. Assuming a linear growth in consumption, the largest growth is mentioned to be +2.5 TWh/year (DNV, 2021). A minimum growth is found to be +0.1 TWh/year (Bruce et al., 2019). This growth is based on the whole of Sweden, combining these growths with the assumption that the SE3 area consumes about 60% of the total energy, the 2035 demand scenarios are presented in Table 1. Creating a

Table 1. The total electricity demand in SE3 based on different scenarios in 2035.

Scenario	SE3 demand in 2035
Minimum	81.2 TWh
Maximum	117.2 TWh

mean demand from these scenarios and adding a deviation from the mean to the different scenarios allows for a comparison between the demand scenarios. The SARIMA forecasting results have been added together to show a total monthly supply up to 2035. How the demand and supply forecasting results compare is shown in Fig. 3.

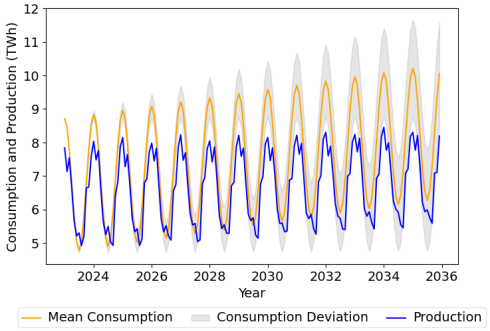


Figure 3. Visualization of the total forecasted power supply and a minimum and maximum demand growth scenario to the year 2035.

From Fig. 3 it can be observed that initially the total monthly supply can match the total monthly demand during the summer months, when the demand is lower compared to the colder winter months. This power shortage during winter months continues, but decreases with an increases supply. The growing supply does not match the mean in growing demand and is increasing more than the supply, increasing the scenarios in which there would be shortages. During high demand periods, the current trends in supply will result in shortages for all scenarios. The forecasting of future demand scenarios relies on literature, each with its own boundary assumptions. The minimum demand value found is lower than the mean historic demand from 2007 to 2022. To maintain current demand levels with the expected electrification of multiple sectors, significant efficiency improvements would be required. The paper justifies this assumption based on expected demand savings in the housing and services sector. For production forecasting, SARIMA models are used, which work best with appropriate statistical values. However, since the goal is to obtain technologically possible values rather than the best forecast, data transformations and tests are not performed. The obtained forecasts align well with historic data for nuclear, solar, and wind power, but are slightly off for hydro and thermal power. The seasonality of these sources matches historic data, but the variation in peaks is not predicted. Improving the forecast for thermal power would require accounting for weather changes as a result of increasing global temperatures, which have uncertainties for the medium to long term. The presented results are within the bounds of current production and technically feasible to continue in the future.

3.2 *Electricity price*

To use a realistic price when determining the payback time of the turbines, the historic day-ahead electricity prices are shown in Fig. 4. The figures shows a decomposition of the price from 2017 to mid-April 2023. The data is extended to show the recent changes in data. The top line in the plot represents the monthly mean day-ahead price. The second line represents the trend, which is the motion of the production over time. If the power source follows a seasonal pattern, this pattern can be observed in the third plot. The scatter plot represents the residual, which is the data that is left after subtracting the trend and the seasonality from the observed data. The residual can be considered noise, the lower the residual, the more stable a source is.

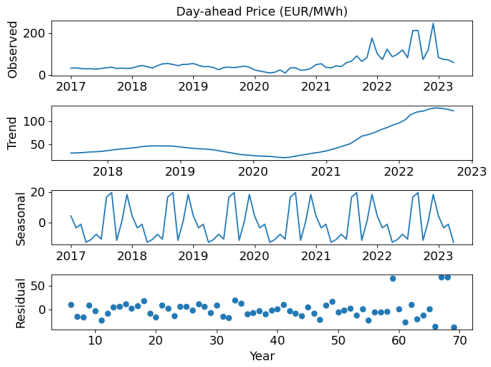


Figure 4. Decomposition of the monthly average electricity price in SE3.

The figure shows a large increase setting in from mid 2020. This rise started at the end of the COVID-19 pandemic and growing international demand (European Council, 2023). The war between Ukraine and Russia, but also heatwaves in the summer of 2022 resulted in a further increase in electricity price (European Council, 2023). The price is also seen to decrease rapidly from 2023. As can be seen from the residual plot, the stability increased from 2022. To neglect the instability and large increased prices, only the months in 2023 are used to determine an average price used later-on in determining the payback time. The seasonality observed in the figure is also neglected and only one value is used. The average price observed in 2023 is 73.81 EUR/MWh and is used for calculating payback times.

3.3 *Wind Analysis*

The weather greatly affects energy consumption, but with the introduction of renewables, production is increasingly influenced by it. To observe the behaviour of the wind around SE3, the data from the stations is analyzed to observe the seasonal differences. The monthly average wind speeds throughout SE3 have been visualized with the help of a Gaussian process regression model. To obtain an understanding of the average monthly wind speeds, the speeds over the region is shown for three separate months in Fig. 5. The wind speeds are gathered from stations with an elevation of 10m above sea level and are transformed to 200m above sea level to improve the clarity of difference in wind speed over land and over sea. With blue representing the lower wind speeds, and red higher wind speeds.

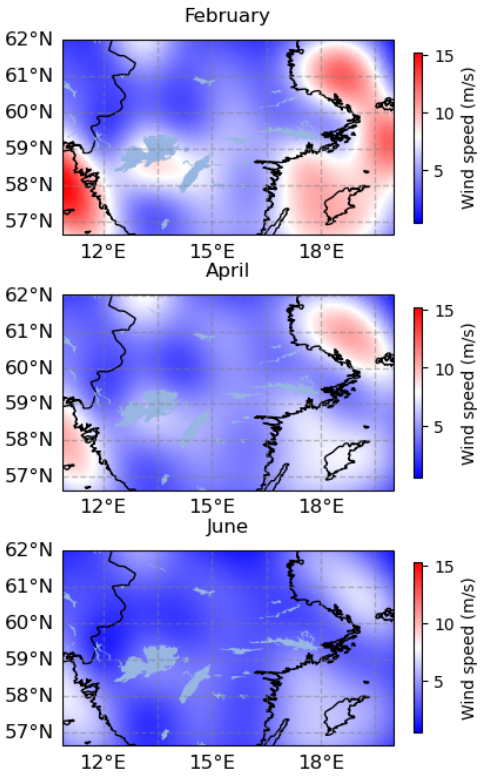


Figure 5. Average monthly wind speeds at 200m above ground level for February, April and June around the SE3 area in 2020.

It is found that the all months have higher wind speeds at offshore areas. The outer northeast and southeast show low wind speeds off the shore. This is due to a lack of stations in that area. Nevertheless, both areas are outside the exclusive economic zone, thus is does not influence the outcome of potential for wind power in SE3. Important to note is the change in scale of the wind speeds for the different months. The average monthly speeds can go up to 15 m/s in February, but also only reach half of that at about 7 m/s in June. The area between the greater lakes in western Sweden also show larger wind speeds. The higher wind speeds are recorded in the months October to March.

3.4 *Wind power distribution*

From the wind speed distribution, the shape and scale parameters can be determined. From this (3) and (4)

can be used to calculate one turbine can generate in a grid cell, after which its capacity factor can be determined. Before the power can be calculated, the wind speeds are transformed to the speed at hub height of the turbine using the surface roughness length of the grid cell.

3.4.1 Current power capacities

The power that can be generated annually throughout SE3 with an onshore turbine having a power capacity of 3.5 MW and a hub height of 100 m and an offshore power capacity of 12.5 MW at 200m and its capacity factor is visualized in Fig. 6.

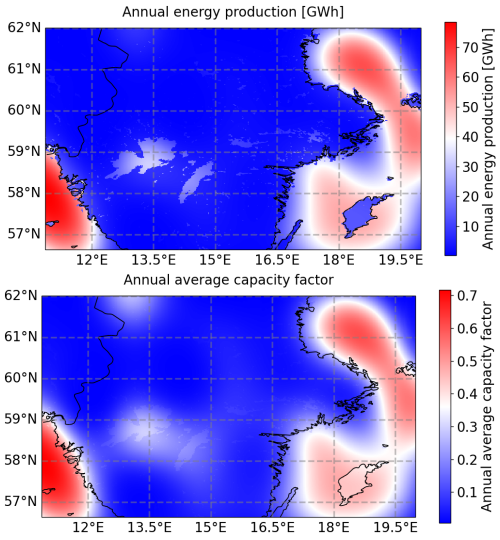


Figure 6. Visualization of the annual energy that can be produced around SE3 with a 3.5 MW onshore and a 12.5 MW offshore turbine and its respective capacity factor at the different sites.

The areas with higher capacity factors correspond to the areas where higher wind speeds are observed in Fig. 5. These areas in turn generate more power.

3.4.2 Future power capacities

How turbine capacities impact the potential is studied by increasing the turbine size to future potential turbine sizes considering technological advancements. With an onshore capacity of 15 MW at 190m and an offshore capacity of 25MW at 200m, the power that can be generated annually around SE3 and the respective capacity factor is visualized in Fig. 7.
The doubling in offshore production is explained by the doubling in capacity factor. The amount of energy that is produced onshore increases more than the increase in capacity factor due to the higher hub heights and thus the availability of higher wind speeds.

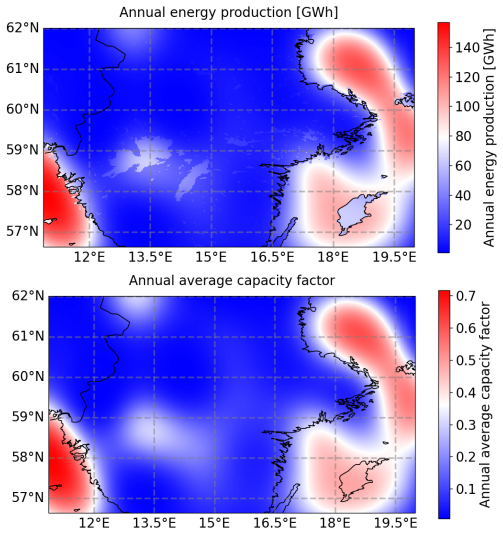


Figure 7. Visualization of the annual energy that can be produced around SE3 with a 15 MW onshore and a 25 MW offshore turbine and its respective capacity factor at the different sites.

3.5 Payback time

Combining the result of the electricity price and the annual power production, a payback time can be calculated. Payback time is defined as the ratio between annual cashflow and total investment cost which is determined as an average of 2021 European costs. The weighted average total installed costs for onshore in Europe is 1623 USD/kW, whereas for offshore it is 2775 USD/kW (IRENA, 2022). Which are assumed to stay the same. Annual operating costs are approximated based on public information from the Swedish Energy Agency. Payback times for all individual cells along the examined design space are calculated.

3.6 Current power capacities

Doing the calculations based on the results from Section 3.4.1, payback times are found to be generally below 10 years for offshore regions, which is visualized in green. Onshore however, a major part has high payback times of more than 20 years, visualized in red in Fig. 8.
Where the pattern observed shows high similarities with the pattern of the capacity factor distribution in Fig. 6. Looking into how much power can be generated in a year, the capacity can be calculated. Doing this for the regions with a payback time below 10 years, a total technical and economical capacity is determined for onshore and offshore turbines. Onshore a capacity of 14GW is found, while for offshore a capacity of 304GW is found.

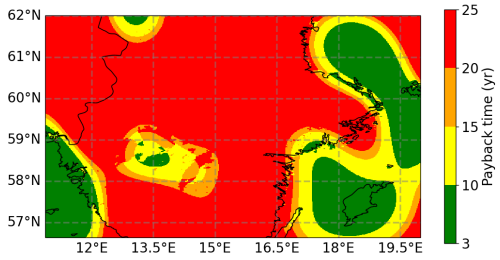


Figure 8. Distribution of payback times around SE3 for onshore and offshore turbines with current turbine capacities.

3.7 Future power capacities

Doing the same based on the results in Section 3.4.2 for the future capacities, a payback time distribution of a single turbine is visualized in Fig. 9.

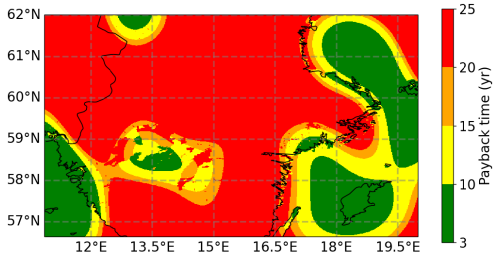


Figure 9. Distribution of payback times around SE3 for onshore and offshore turbines with turbine capacities possible in 2035.

The study identifies improved onshore and offshore capacities of 83 GW and 608 GW, respectively, with a payback time below 10 years. Local prices and changes in 2035 need to be considered for better accuracy. Power production and capacity factors show linear trends with higher capacity factors resulting in lower payback times. However, the calculations focus on individual turbines and do not account for the efficiency decrease in wind farms.

3.8 2035 power production

To put the results from Section 3.7 into perspective, the capacity is compared with the forecasted demand and supply in 2035. The maximum gap in demand and supply is found to be 33.7 TWh. Additional 3.85 GW wind power capacity would be needed to meet the annual demand. In a wind farm configuration, the turbines are assumed to lose 20% of their capacity in the south of Sweden (Holttinen, 2005). Assuming an offshore wind farm of 75 turbines in locations with a payback time below 10 years, where the capacity factor is averaged to 0.55, a farm can produce 7.5 TWh per year. Which means 4.5 of such farms can add to the total annual energy demand in 2035. Hence, the technically potential energy generated is sufficient to meet the annual energy demand in 2035 when considering payback times. However, the area studied is not fully within the exclusive economic zone and does exclude governmental restricted areas for wind farm sites. Next to this, the seasonality of energy production and the impact on the grid need further study. Factors such as load factor and grid losses should be considered to evaluate the effect of additional turbines on the grid. The calculated amount of farms is based on balancing annual consumption. As can be observed in Figs. 2 & 5, there is a seasonal dependency in available wind power. Renewable energy actually has hourly deviations which are difficult to predict, especially for long-term ahead. The hourly balancing between demand and supply is crucial to operate a stable grid. For proper balancing, the surpluses need to be well managed. Storing the surpluses is a crucial part to be able to supplement the shortages in production at other times.

4 Conclusion

The goal of this research was to study the technical potential capacity for wind turbines in the SE3 area. Based on the analysis performed here, it is found that the potential for supplementary wind power in the area is large. With the sustainability goal of reaching 100% in 2040, the role of wind power is important to be studied. This paper contributes to the knowledge of wind power in the area by comparing the current trends in individual power supply sources with literature values for future demand and observing that if the current trend is continued, the supply in 2035 can barely meet the demand. This means that a growth in supply is likely necessary in order to reduce the costs of electricity in SE3, as has been the goal of the introduction of different bidding in Sweden. Having observed the need for additional power in 2035, the role wind power can play has been studied. Both current turbine capacities and expected future turbine capacities can be used to locally produce additional renewable power considering initial economical factors and reach the sustainability goal. The tool made in this research allows for focusing on a specific area and performing basic economical calculations, which can be expanded beyond the payback time by adding additional parameters.

Acknowledgement

The authors would like to express their gratitude to mr. B. Fält from Svenska Kraftnät for his discussion about the historic demand and supply data and increase our understanding of the system and improve the data analysis progress. This work has been partially financed by the project GEFWIN, funded by Vinnova, under the umbrella of Eurostars/Eureka.

References

Armelius, H. (2022, Sep). *Electricity shortage.* https://www.ekonomifakta.se/Fakta/Energi/Energibalans-i-Sverige/elbrist/.

Back, A. (2020). *Footprints of an invisible population: second-home tourism and its heterogeneous impacts on municipal planning and housing markets in sweden* (Unpublished doctoral dissertation). Umeå universitet.

Badger, J., Davis, N., Hahmann, A., Hansen, B. O., Mortensen, N. G., Olsen, B. T., … Volker, P. (2023, Apr). *Global wind atlas.* https://globalwindatlas.info/en/. (Data extracted on 23-03-2023)

Boyle, G. (1996). *Renewable energy: power for a sustainable future* (Vol. 2). Oxford University Press.

Bruce, J., Krönert, F., Obel, F., Yuen, K., Wiesner, E., Dyab, L., … others (2019, Jun). Färdplan fossilfri el–analysunderlag-en analys av scenarier med en kraftigt ökad elanvändning. *Sweco, Profu, IVL, Augusti*. (Translated using Google Translate)

Burton, T., Jenkins, N., Sharpe, D., & Bossanyi, E. (2011). *Wind energy handbook*. John Wiley & Sons.

DNV. (2021, May). *Study on opportunities and barriers to electrification in the nordic region.* https://www.energiforetagen.se/globalassets/dokument/. (Accessed on 13-02-2023)

Energimyndigheten. (2018, Oct). *Strategi för hållbar vindkraftsutbyggnad miljömålsrådsåtgärd 2018.* http://www.energimyndigheten.se/globalassets/fornybart/framjande-av-vindkraft/uppdragsplan-1.0.pdf. (Accessed on 05-12-2022. Trans. Google Translate)

ENTSO-E. (2023, Apr). *Day-ahead prices.* https://transparency.entsoe.eu/transmission-domain/r2/dayAheadPrices. (Data extracted on 23-01-2023)

eSett. (2023). *Production.* https://opendata.esett.com/production. (Accessed on 14-02-2023)

European Council. (2023, Mar). *Infographic - energy price rise since 2021.* https://www.consilium.europa.eu/en/infographics/energy-prices-2021/. (Data extracted on 28-01-2023)

Gawel, E., Lehmann, P., Purkus, A., Söderholm, P., & Strunz, S. (2017). *Political economy of safe-guarding security of supply with high shares of renewables*. Retrieved from https://energiforskmedia.blob.core.windows.net/media/23204/political-economy-of-safe-guarding-security-of-supply-with-high-shares-of-renewables-energiforskrapport-2017-441.pdf

Holttinen, H. (2005). Hourly wind power variations in the nordic countries. *Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology*, *8*(2), 173–195.

IEA. (2019). *Energy policies of iea countries: Sweden 2019 review*. IEA, Paris. https://www.iea.org/reports/energy-policies-of-iea-countries-sweden-2019-review.

IRENA. (2022). *Renewable power generation costs in 2021*.

JB Sustainable Approach AB. (2019, Aug). *White paper: Solar energy in sweden rapid market growth – consolidation to follow* (Tech. Rep.).

Råberger, C., & Vingmarker, V. (2019). *Klimatstrategi för västmanlands län* (G. Translate, Trans.). Retrieved from https://catalog.lansstyrelsen.se/store/23/resource/DU_2019_08 (Trans. Google Translate)

Shu, Z., & Jesson, M. (2021). Estimation of weibull parameters for wind energy analysis across the uk. *Journal of Renewable and Sustainable Energy*, *13*(2), 023303.

Statens Energimyndighet. (2022, May). *Energiindikatorer 2022 uppföljning av sveriges energipolitiska mål.* https://energimyndigheten.a-w2m.se/Home.mvc?ResourceId=206547. (Accessed on 05-12-2022. Trans. Google Translate)

Svenska Kraftnät. (2021, May). *Kraftbalansen på den svenska elmarknaden, rapport 2021 (ärendenr 2021/1042).* (Translated using Google Translate)

Svenska Kraftnät. (2007-2022). *Elstatistik.* https://www.svk.se/om-kraftsystemet/kraftsystemdata/elstatistik/.

Svenska Kraftnät. (2022). *The control room.* https://www.svk.se/en/national-grid/the-control-room/. (Accessed on 26-04-2023)

SWEA. (2021, Jan). *Roadmap 2040.* https://swedishwindenergy.com/wp-content/uploads/2021/01/Roadmap-2040-ENG-rev-2020.pdf. (Accesed on 05-12-2022)

SWEA. (2022, Oct). *Statistics and forecast wind power sweden-q3 2022.* `https://swedishwindenergy.com/statistics`. (Accessed on 01-06-2023)

Simulation of blue hydrogen production by natural gas in the North Sea

Chidapha Deeraksa[a,*], Nora Cecilie Ivarsdatter Skau Furuvik[a] Britt Margrethe Emilie Moldestad[a]

[a] *Department of Process, Energy and Environmental Technology, University of South-Eastern Norway*

Chidapha.Deeraksa@usn.no
britt.moldestad@usn.no

Abstract

Hydrogen is an efficient energy carrier and an important contribution to sustainable energy development. Hydrogen can be produced based on different methods and on different raw materials. Blue hydrogen is hydrogen produced from natural gas via a steam-methane reformer with subsequent carbon capture and storage. The CO_2 from the process can be stored in matured oil and gas fields or in an aquifer.

This paper studies the potential of producing blue hydrogen from methane from the Troll gas field on the Norwegian continental shelf. The production rate of methane from the Troll field is predicted and based on the calculated methane production the steam-methane reformation process is modelled and simulated. The model includes the required steps to convert natural gas into hydrogen and CO_2 and further to catch the CO_2. The volume of captured CO_2 per m^3 of produced hydrogen is calculated. Production of blue hydrogen also includes storage of CO_2, and the required storage capacity is calculated.

The purpose of this paper was to investigate whether blue hydrogen produced by natural gas from the Troll field is an alternative to reducing CO_2 emissions to reach the climate target. The simulation was performed with Aspen HYSYS 12 and the calculation on how much CO_2 must be stored and the storage capacity needed were performed manually. The mass of CO_2 resulting from the conversion of about 2400 tons natural gas/h to blue hydrogen and CO_2 at the Troll field is 5600 tons CO_2/hour or 49 megatons CO_2/year. The produced hydrogen had a purity of 95%. The predicted storage capacity for CO_2 at the Troll field is found to be 136 megatons. A profitability analysis is performed and the results are promissing.

1. Introduction

The main cause of climate change is the emissions of greenhouse gases such as carbon dioxide (CO_2), methane (CO_4) and nitrous oxide (N_2O) (European commission, 2023). However, the CO_2 emission is the largest contributor to global warming. The average temperature on earth has increased gradually by at least 1.1°C since 1880 and most of the warming has appeared since 1975 (Nasa.gov, 2023). If the increasing temperature is not limited to 1.5°C but reaches 2°C, serious consequences will arise. For instance, there will be less insect death which lead to less production of rice, corn and other food products. Likewise, there will be less fish in the seas since approximately 70% to 90% of coral reefs will die. Moreover, by 2050 over 300 million people will be affected by the rising seas. Reducing global greenhouse gas emissions will activate the temperature to fall back down and then the climate system probably stabilizes again (NPR, 2021).

For limiting the global temperature rise to 1.5 °C, the Paris Agreement has laid the foundation for the world to cut the greenhouse gas emissions by 45% by 2030 compared to the 2010 level (United Nation Climate Change, 2022) Norway was among 175 countries that have committed to the agreement. Norway is now further increasing the target by submitting to the UN ahead of the UN Climate Change Conference (COP27) in Egypt. The new target is to reduce emissions at least 55 % compared to the 1990 level by 2030 (Government.no, 2022)

The emissions that are focused on in this paper are CO_2 emissions from oil and gas extraction, which have a quantity of about 12.2 million tons CO_2 equivalents. It amounts to approx. 32% of the Norwegian CO_2-emissions (Statistics, Norway, 2023), (Mustafa *et al*., 2016). There are projects that will reduce the greenhouse gases in Norway such as Carbon Capture and Storage (CCS) at Norcem and Klemetsrud and renewable energy like hydropower, wind power, solar energy, and hydrogen.

Hydrogen is a highly efficient energy carrier and is emitting only water vapor when reacting with oxygen. Hydrogen is used in fields like petroleum refining, ammonia production, methanol production, power generation, and transportation. (Energy

efficiency&renewable energy, 2023). To meet the Paris Agreement and the COP27 targets, it is required to develop cost-effective low CO_2 emission hydrogen production technologies. Hydrogen can be produced based on different methods and different raw materials, and is named black, grey, green, and blue hydrogen. Black and gray hydrogen is produced using coal and natural gas respectively, and without any CO_2 capture and storage (CCS). Green hydrogen is hydrogen produced from renewable resources. Blue hydrogen is hydrogen produced from natural gas in a gas reformer with CCS (National**grid**, 2023). Steam methane reforming, autothermal reforming, and natural gas decomposition are technologies used to produce blue hydrogen from natural gas. This paper focuses on steam methane reforming with CCS.

Norway is one of the largest exporters of natural gas in the world. Norway was in the third place in 2021 and covered approximately 23% of the gas demand in EU and United Kingdom. Norway exported about 122 billion Sm^3 of natural gas in 2022 and roughly 60% of Norway's natural gas resources have not yet been produced (Norwegian Petroleum, 2022). A large part of the natural gas production in Norway comes from the Troll field, which is the largest field in the North Sea (Equinor, 2023).

To achieve the climate target Norway must shut down some of its old oil and gas fields prematurely unless they can use carbon-free technologies to cut down their emissions (Reuters, 2022). This paper studies the potential of producing blue hydrogen from the Troll field by using the Aspen HYSYS V12 software. The paper covers a detail analysis to determine whether this process is an economically and environmentally friendly way of handling natural gas. In other words, this paper investigates whether the existing gas fields can be used to convert the natural gas to a clean, reliable and affordable hydrogen, instead of shutting them down. Norwegian Petroleum's website indicates that Troll produced 37.36 million Sm^3 o.e. natural gas in 2021 and this paper assumes that natural gas from the Troll field contains 92.74 vol.% Methane (CH_4), 1.83 vol.% CO_2, 0.0045 vol.% Nitrogen (N_2) 4.07 vol.% Ethane(C_2H_6) and 0.91 vol.% Propane (C_3H_8) (Aromada and Kvamme, 2019), (Norwegian Petroleum Troll, 2021).

2. Methodology

Gas reforming using steam is the most common and cheapest method of producing hydrogen and therefore over 95% of the world's production of hydrogen is based on the steam methane reforming process (SMR). There are process plants that have a production capacity of anywhere from 1 to 100 tons of hydrogen per hour. (Rapier, 2020), (Gupta, 2008)

2.1 Modelling of the steam methane reforming process

Gas reforming is a chemical process where natural gas is reacted with steam using a catalyst at quite high temperature to produce carbon monoxide (CO) and hydrogen (H_2) (Gupta, 2008).

The process starts with the natural gas being pre-treated, where organic sulphur compounds (thiols) are converted to hydrogen sulfide (H_2S), using a catalyst such as porous aluminium filled with cobalt (Co), nickel (Ni), molybdenum (Mo) and wolfram (W). The H_2S is removed from the stream by using a catalyst consisting of zinc oxide (ZnO). This step is used for preventing sulphur from polluting downstream catalysts (Gupta, 2008).

The sulphur-free natural gas is processed further in pre-reforming process which is the process that takes place before the reformer. In this process, the larger hydrocarbons are broken down into CH_4, CO_x and H_2 in an adiabatic reactor at a temperature around 300-525°C. The catalyst mass in the reactor consists of aluminium containing nickel. The advantage of pre-reforming is that the plant can run a natural gas feed stream with varying contents of larger hydrocarbons, and the steam-carbon ratio is reduced so that the effect of the plant is increased. The chemical reaction equation for hydrocarbons is (Gupta, 2008):

$$C_nH_m + nH_2O \rightleftharpoons nCO + \left(\frac{m+2n}{2}\right)H_2 \quad \text{(Rn1)}$$

The natural gas, which now consists of mostly CH_4, CO_x and H_2 is mixed with steam having a pressure of approx. 20 – 26 bar. This mixture is heated before being fed to a catalytic reforming reactor which contains tubes filled with nickel catalyst. In the reformer, methane reacts with water and is converted to CO and H_2 according to the major steam reforming reaction (Rn2) and is then converted to CO_2 and H_2 according to the steam reforming reaction (Rn3) (Gupta, 2008).

$$CH_4 + H_2O \rightleftharpoons CO + 3H_2 \quad \Delta H^0_{298} = +206 kJ/mol \quad \text{(Rn2)}$$

$$CH_4 + 2H_2O \rightleftharpoons CO_2 + 4H_2 \quad \Delta H^0_{298} = +165 kJ/mol \quad \text{(Rn3)}$$

The reactions are endothermic, which means that the reactions absorb energy from the surroundings, and the enthalpy change from the reaction requires approx. 206 kJ/mol for Rn2, and 165 kJ/mol for Rn3. These energies are supplied by burning some of the natural gas, but electricity (EL) should also be considered. The temperature required is between 700°C and 950°C. Low pressures are preferred for the reactions, but because most industries require H_2 at a pressure of at least 20 bar the reformer is run at a pressure around 20 to 26 bar. High pressures allow a more compact reactor design, increased reactor output, and reduced material costs. According to the stoichiometry in the reactions Rn2 and Rn3, the ratio between methane and steam is 1:1 and 1:2 on a molar basis. In practice, excess steam is used to prevent carbon build-up, hence the ratio1:2 for methane and

steam is chosen in the HYSYS simulations (Gupta, 2008).

The steam reforming reactions Rn4 and Rn5 for C_2H_6 and C_3H_8 were not included in the stochiometric reactions in HYSYS V12.

$$C_2H_6 + 2H_2O \rightarrow 2CO + 5H_2 \quad \text{(Rn4)}$$

$$C_3H_8 + 3H_2O \rightarrow 3CO + 7H_2 \quad \text{(Rn5)}$$

The stream from the reforming stage consists of H_2, CO, CO_2, water vapor and a small proportion of CH_4 that has not been reformed. This stream is processed further in the water-gas shift reactors. Here, CO reacts with steam over a catalytic bed and produces H_2 and CO_2 as seen in reaction (Rn6). The lower temperature with respect to the reformer is needed for this reaction, since it is thermodynamically preferred at low temperatures. This is an exothermic reaction and emits 41.2 kJ/mol (Gupta, 2008).

$$CO + H_2O \rightleftharpoons H_2 + CO_2 \quad \Delta H^0_{298} = -41.2\ kJ/mol \quad \text{(Rn6)}$$

The excess water is separated from the gas stream by using a separator with low temperature. Here, the water vapor is condensed and leaves the separator in the gas stream. The H_2 and the CO_2 flows are separated by capturing the CO_2 with monoethanol-amine (MEA) in an absorption tower. The CO_2 will be stored in the reservoir or an aquifer, and H_2 is further sent to a purification process. The remaining CO_2 and CO are removed in a final step called methanisation, where these components are converted into CH_4 as shown in Rn7 and Rn8 (Gupta, 2008).

$$CO + 3\,H_2 \rightleftharpoons CH_4 + H_2O \quad \text{(Rn7)}$$

$$CO_2 + 4\,H_2 \rightleftharpoons CH_4 + 2H_2O \quad \text{(Rn8)}$$

The stream can be processed further in an activated carbon adsorber to separate CH_4 from the H_2 product. The CH_4 stream is then recycled to the reformer. If the final stream has only H_2 as the product, the recirculation can be skipped. Fig. 1 shows the hydrogen reforming process using the Aspen HYSYS V12. The model includes the required steps to convert natural gas into hydrogen and CO_2. Then the CO_2 is captured with amine-based solution and converting a small proportion of CO back to CH_4. In the simulation, the main processes occur in the reformer, two water-gas shift reactors, a separator, a CO_2 absorber and a methanator. However, there are some processes that are not included in the simulations such as the pre-treatment process, the pre-reforming process, and the activated carbon adsorption process.

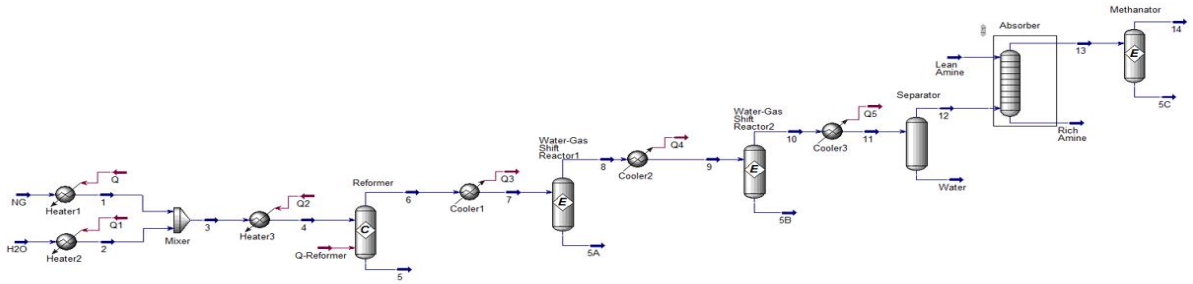


Figure 1: Steam reforming process Aspen HYSYS V12 simulation

2.2. CO_2 storage capacity

A technology for reducing CO_2 emissions is called geological carbon sequestration (GCS) (Lackner, 2003), (Schrag, 2007). Deep saline aquifers have large storage capacity and is therefore well suitable for GCS (Bachu, 2003). However, to ensure a safe utilization of CGS to a particular aquifer, an accurate calculation of the storage capacity of the aquifer is required. Different models can be used to calculate the storage capacity of CO_2 in an aquifer. In this study a model developed by Szulczewski and Juanes (Szulczewski and Juanes, 2009) is used to calculate the mass of trapped CO_2. The model is simple and robust and includes some assumptions to be made. The reservoir is assumed to be horizontal, homogeneous, and isotropic. Other assumptions are that the injected CO_2 follows the direction of the groundwater and that the viscosities and densities of the fluids are constant. It is also assumed that there is a sharp interface between the CO_2 plume and the brine. The storage capacity, C, is calculated from:

$$C = \left[\frac{2M\Gamma^2(1-S_{cw})}{\Gamma^2 + (2-\Gamma)(1-M+M\Gamma)}\right]\rho_{CO_2}\varphi HWL_{tot} \quad (1)$$

where M is the mobility ratio, Γ is the trapping coefficient, S_{cw} is the connate water saturation, ρ_{CO_2} is the density of CO_2, φ is the porosity, H is the thickness of the sandstone, W is the length of the injection array, and L_{tot} is the total length of the simulated reservoir. The storage efficiency, which is the term in brackets in Equation (1), relates the total pore volume to the volume of trapped CO_2. The mobility ratio is expressed as:

$$M = \frac{1/\mu_w}{k^*_{rg}/\mu_{CO2}} \quad (2)$$

where μ_w and μ_{CO2} are the viscosity of brine and CO_2, and k^*_{rCO2} is the endpoint relative permeability of supercritical CO_2. The trapping coefficient, Γ, is defined as:

$$\Gamma = \frac{S_{rCO2}}{1-S_{cw}} \quad (3)$$

where S_{rCO2} is the residual saturation of CO_2 and S_{cw} is the connate brine saturation. The CO_2 storage model developed by (Szulczewski and Juanes, 2009) also includes an equation for the CO_2 footprint. The equation calculates how far the CO_2 plume migrates away from the injection array when it is completely trapped. The distance is expressed as:

$$L_{max} = \left[\frac{(2-\Gamma)(1-M(1-\Gamma))}{(2-\Gamma)\big(1-M(1-\Gamma)\big)+\Gamma^2}\right] L_{tot} \quad (4)$$

The CO_2 footprint is illustrated in Fig. 2. The injection footprint has a length, L_{inj}, and is defined as the distance the CO_2 plume is moving during the injection period and is expressed by:

$$L_{inj} = L_{tot} - L_{max} \quad (5)$$

The injection footprint is marked with darker blue in Fig. 2. The light blue area in the figure presents the trapped CO_2 footprint and has an extent L_{max}. The blue arrays show the groundwater flow direction.

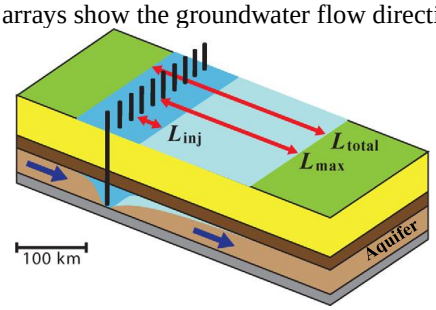


Figure 2: Injection and trapped CO_2 footprints

3. Results and discussion

3.1. Steam methane reforming process simulation

The Peng Robinson equation of state was selected for the simulations due to the types of gas components, chemical reactions and equipment used. A conversion reactor was selected as the reformer and an equilibrium reactor was selected as the Water-Gas Shift reactors and the Methanator. The methane to steam ratio, the pressure (P), and the temperature (T) were adjusted following the methodology in order to achieve the optimum results.

The calculation of energies used for both heating and cooling in the system is done in term of electricity. The heating and cooling duties required to operate at optimal conditions were computed by the HYSYS V12 simulator by adjusting the temperatures in and out of the reactors, the coolers, and the heater. Based on this, the electricity cost was calculated in Excel, by assuming that the electricity price was 0.5 NOK per kWh. The obtained results are shown in Tab. 1.

Table 1: Heat duties of reformer, heaters and coolers from HYSYS simulations. Electricity costs using 0.5 NOK/kWh

El for Heating		**El for cooling**	
Energy Stream	Duty [kW]	Energy Stream	Duty [kW]
Q	405900	Q3	3170000
Q1	511300	Q4	1119000
Q2	1329000	Q5	1700000
Q-Reformer	7297000		
Total:	9543200	Total:	5989000
Cost:	$4.18*10^{10}$ [NOK/year]	Cost:	$2.62*10^{10}$ [NOK/year]

The gas components in and out of the reformer, the water-gas shift reactors, the separator, the absorber and the methanator were likewise computed by the HYSYS simulator, where 85% efficiency was chosen for the reformer. Tab. 2 shows the obtained results where the produced CH_4 was converted into 69.54% H_2 after going through the reformer and the two water-gas shift reactors. The concentration of H_2 increased to 95.33% after passing the purification and methanation process.

Table 2: HYSYS simulation output data for the reformer, Water-Gas Shift reactors, separator, absorber and methanator

	Feed		**Reformer** $\eta = 0.85$	**Water-Gas Shift reactors**		**Separator**	**Absorber**	**Methanator**	
Component	**Ng**		**Outlet gas 6**	**Outlet gas 8**	**Outlet gas 10**	**Outlet gas 12**	**Outlet gas 13**	**Outlet gas 14**	
	Molar percent (%)	Molar flow rate [kmol/h]	Molar percent (%)	Molar percent (%)	Molar percent (%)	Molar percent (%)	Molar percent (%)	Molar percent (%)	Molar flow rate [kmol/h]
CH_4	0.9274	148940	0.0309	0.0309	0.0309	0.0335	0.0331	0.0394	19739
CO_2	0.0183	2938	0.0915	0.1490	0.1747	0.1894	0.0001	0.0001	39
CO	-	-	0.0875	0.0300	0.0043	0.0047	0.0058	-	-
H_2	-	-	0.6122	0.6697	0.6954	0.7542	0.9596	0.9533	477946
H_2O	-	-	0.2659	0.1084	0.0827	0.0052	-	0.0058	2932
N_2	0.0045	722	0.0010	0.0010	0.0010	0.0011	0.0013	0.0014	679
C_2H_6	0.0407	6536	0.0090	0.0090	0.0090	0.0098	-	-	16
C_3H_8	0.0091	1461	0.0020	0.0020	0.0020	0.0022	-	-	0.0002

3.2. Net profit calculations.

Since this paper assumes that the natural gas from the Troll field contains 92.74 vol% CH_4 and that Troll produced 37.36 million Sm^3 o.e. natural gas in 2022, then the mass flow rate of CH_4 is calculated to be approximately 2400 tons/h or 148940 kmol/h. After the gas stream is passing through the steam reforming process the concentration of the H_2 in the final product is 95.33 % which corresponds to about 960 tons/h or 8444000 tons/year.

Different price ranges in the market provide different incomes from selling the blue hydrogen. Tab. 3 shows the estimated income based on various hydrogen prices where the highest income from selling hydrogen is $1.235 \cdot 10^{12}$ NOK/year when the sale price is 195 NOK/kg. According to (glpautogas.info, 2023) the average price of hydrogen in Norway in August 2023 is 195 NOK/kg. This is the price for the customers at the hydrogen refueling stations. The price is including 25% VAT, which means that the real income is 146.25 NOK/kg.

Table 3: Income from selling H_2

Amount H_2		Without VAT 25%	
$8.444 \cdot 10^9$ [kg/year]	Sale Price [NOK/kg]	Real price [NOK/kg]	Income [NOK/year]
Alternative 1	159	119.25	$1.007 \cdot 10^{12}$
Alternative 2	195	146.25	$1.235 \cdot 10^{12}$

The other income comes from using CCS in the process. Hence, there is a price that emitters must pay per tonne of CO_2 emission. This price is the summation of the carbon tax and the emission trading system set by the government. Hence, carbon emissions have a cost, and reducing CO_2 in the process will reduce this cost (avoided cost). This reduction in costs can count as an income to the project (Norwegian Petroleum, emissions, 2022). In Norway, the companies pay approximately 1100 NOK/ton for their CO_2 emissions (Norwegian Petroleum, emissions, 2022). The process simulation results show that the blue hydrogen process at the Troll field can help to reduce the greenhouse gases by almost 49 megaton CO_2 per year (5600 tons CO_2/h) which corresponds to $5.39 \cdot 10^{10}$ NOK/year. The net profits which is the sum of income from selling H_2 and the profit from reducing CO_2 emission are shown in Fig. 3, where the income level is between $1.06 \cdot 10^{12}$ to $1.29 \cdot 10^{12}$ NOK/year.

3.2. Calculation for utilities cost with various electricity price in Norway.

There is uncertainty related to electricity price in Norway since the price is higher in the winter and lower in the summer and changing all day. Therefore, the electricity cost was calculated with a considerable range of electricity prices (0.5 NOK /kWh, 1NOK /kWh and 1.50 NOK/kWh) to cover the large variations. The obtained results are shown in Fig. 4.

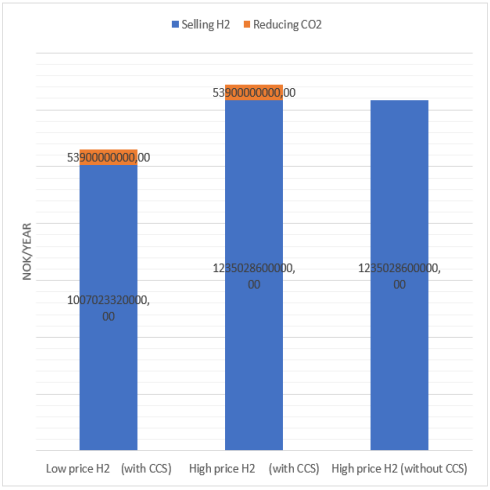


Figure 3: Net income from the H_2 production.

Figure 4: Utilities cost for various electricity prices in Norway

3.3. Production costs and operating costs for hydrogen production and CCS.

There are costs related to production and operation for hydrogen production and CCS. Other costs are maintenance and transportation costs that will affect the net profit of the project. In order to be able to calculate the financial impacts of converting natural gas to H_2, it is necessary to take all negative and positive cash flows into account.

However, there are some uncertainties about these values, and therefore only the estimation of the production cost will be present in this section. As seen in Tab. 5 the production cost varies from $1.10 \cdot 10^{11}$ to $2.29 \cdot 10^{11}$ NOK/year when the H_2 production is $8.445 \cdot 10^9$ kg/year. The exchange rate from US$ to NOK is used as 10.64 NOK/US$. Alt. 1 in Tab. 5 is based on data from a Norwegian report (Klimastiftelsen, 2021) and the estimated production cost includes CCS. Alt.2, Alt.3 and Alt. 4 are presented by (Oni *et al.*, 2022). All the alternatives are based on production of hydrogen from steam methane reforming.

Table 5: Estimates of H_2 production costs. The (Bjartnes *et al*., 2021), (Oni *et al*., 2022).

Amount H_2 $8.445 \cdot 10^9$ [kg H_2/year]	H_2 production price rate [NOK/kgH_2]	Production costs [NOK/year]
Alt.1(CCS)	17.02	$1.44 \cdot 10^{11}$
Alt.2(0% CCS)	12.98	$1.10 \cdot 10^{11}$
Alt.3(52% CCS)	17.66	$1.49 \cdot 10^{11}$
Alt.4(85% CCS)	27,13	$2.29 \cdot 10^{11}$

3.4. Profitability analysis

Gas reforming requires energy, which results in lower profitability, but when using CCS, savings from climate taxes can be greater than the expenses. There are some uncertainties regarding these values including some concerns related to simplifying the gas composition of natural gas, which have an impact on the economic perspective. Also, the amount of H_2 produced and the amount of captured CO_2 will be important factors when it comes to assessing the profitability.

There are many factors that can affect the profitability, including the transportation and the price of hydrogen. There are other costs that are not included in this study, such as maintenance cost, deprecitation of investment cost, equipment cost and installation cost for the blue hydrogen process. However, based on the assumptions and calculations that have been made, the profit is large. This means there are good opportunities to produce blue hydrogen from natural gas from the Troll field. For the best-case scenario, the earning after paying the utility and the production costs is $1.11 \cdot 10^{12}$ NOK/year and for the worst-case scenario the earning is $5.74 \cdot 10^{11}$ NOK/year.

3.5 CO_2 storage capacity at the Troll field

The storage capacity at the Troll field is calculated based on the Szulczewski and Juanes model (Szulczewski and Juanes, 2009) and is compared to the CO_2 production from the blue hydrogen process.

The density and viscosity of supercritical CO_2 and brine are calculated based on the temperature and pressure at the Troll field which is given as 60 °C and 100 bar, respectively. The thickness of the sandstone (H), the porosity (φ), the total extent of the CO_2 plume (L_{total}) and the length of the injection formation (W) have been chosen based on older available data from the Troll field. The input parameters for the storage calculations are given in Tab. 6. The calculated storage parameters are presented in Tab. 7.

Table 6: Input parameters for calculating CO_2 storage capacity.

Parameter	
ρ_{CO2}	290 kg/m^3
μ_{CO2}	$2.374 \cdot 10^{-5}$ Pa·s
μ_W	0.00046 Pa·s
S_{rCO2}	0.3
S_{cw}	0.3
φ	0.27
W	40 000 m
H	30 m
L_{tot}	100 000 m
k^*_{rCO2}	0.55

Table 7: Results from the storage capacity calculations.

Parameter	
M	0.0938
Γ	0.4286
E	1.44%
C	$1.36 \cdot 10^{11}$kg CO_2
L_{max}	89 000 m
L_{inj}	11 000 m

It was not possible to find data for the dimensions of the aquifer under the Troll field, and the storage capacity is therefore calculated based on assumed H, W and L_{tot}. The aquifer under the Troll gas field is most properly much larger, and the calculated storage capasity is highly underpredicted.

The mass of CO_2 resulting from the conversion of 2400 tons natural gas/h to blue hydrogen and CO_2 at the Troll field is 5600 tons CO_2/h or 49 megatons CO_2/year. The calculated storage capacity for CO_2 at the Troll field is found to be 136 megatons. This gives a perspective on the required storage space and the potential for CO_2 storage at the Troll field.

4. Conclusion

The purpose of this paper is to investigate whether blue hydrogen produced by natural gas from the Troll field is an alternative to reducing CO_2 emissions and thereby contribute to reach the climate target. The prosess of converting naturalgas to blue hydrogen is modelled, and simulations were performed using Aspen Hysys 12. The model includes the required steps to convert natural gas into hydrogen and CO_2 and further to catch the CO_2. Conversion of about 2400 tons natural gas/h gave 960 tons/h of blue hydrogen with a purity 95%, and 5600 tons /h of CO_2 (49 megatons CO_2/year).

The predicted storage capacity for CO_2 at the assumed Troll field is found to be 136 megatons. However, the aquifer under the Troll gas field are most probable much larger and have a much higher storage capasity than predicted here.

There are good opportunities for blue hydrogen production from natural gas fields in the North Sea. The profit is calculated and the results are promissing.

Acknowledgment

A part of this paper is based on the Bachelor thesis: “Production of blue hydrogen from natural gas fields in the North Sea.” Special thanks to the authors of the thesis Kjetil Kolstad and Ivar André Tors.

References

Aromada Solomon, Kvamme Bjørn (2023) Researchgate Article, 'Gas components'. Accessed: 14.06.2023. Available: https://www.researchgate.net/publication/332415203_Impacts_of_CO_2_and_H_2_S_on_the_risk_of_hydrate_formation_during_pipeline_transport_of_natural_gas/figures?lo=1

Albara Mustafa, Rajnish Kaur Calay, M. Y. Mustafa, and Subhashis Das (2016) ‘Smart production of biofuel for small communities’, *International Symposium on Small-scale Intelligent Manufacturing, Systems*, https://www.researchgate.net/publication/312037541_Smart_production_of_biofuel_for_small_communities/figures?lo=1

Bachu, S. (2003) ‘Screening and ranking of sedimentary basins’, *Environ. Geol.* 44 (2003) pp. 277-289.

Bjartnes, A., Hirth, M. L., Michelsen, N. H. P., Skaugen, H., Ursin, L. (2021) ‘Hydrogen som klimaløsning’, Norsk Klimastiftelse, (3) 2021. https://klimastiftelsen.no/publikasjoner/hydrogen-som-klimalosning/

Earth-Observatory, Nasa.gov, 'World of Change: Global Temperatures'. Accessed: 14.06.2023. Available: https://earthobservatory.nasa.gov/world-of-change/global-temperatures

Energy Efficiency & Renewable Energy, 'Hydrogen: A Clean, Flexible Energy Carrier'. Accessed: 14.06.2023. Available: https://www.energy.gov/eere/articles/hydrogen-clean-flexible-energy-carrier

Europian commision, 'Causes of climate change'. Accessed: 14.06.2023. Available: https://climate.ec.europa.eu/climate-change/causes-climate-change_en

Equinor, ‘The Troll field and the Troll A, B and C platform’. Available: https://www.equinor.com/energy/troll

Glpautogas.info, 'Hydrogen fuel price in Norway'. Accessed: 14.06.2023. Available: https://www.glpautogas.info/en/hydrogen-sale-price-norway.html

Government.no, ‘Norway’s new climate target: emissions to be cut by at least 55 %’, Accessed: 14.06.2023, Available: https://www.regjeringen.no/en/aktuelt/norways-new-climate-target-emissions-to-be-cut-by-at-least-55-/id2944876/

Gupta, R. B.(2008) ‘Hydrogen Fuel: Production, Transport and Storage’, CRC Press.

Lackner, K. S. (2003) ‘A guide to CO2 sequestration’, *Science* 300, pp. 1677-1678.

Nationalgrid, 'The hydrogen colour spectrum'. Accessed: 14.06.2023. Available: https://www.nationalgrid.com /stories/energy- explained/hydrogen-colour-spectrum

Norwegian Petroleum, 'Emissions to air'. Accessed: 04.07.2023. Available: https://www.norskpetroleum.no /en/environment-and-technology/emissions-to-air/

Norwegian Petroleum, 'Export of oil and gas'. Accessed: 14.06.2023. Available: https://www.norskpetroleum.no/en /production-and-exports/exports-of-oil-and-gas/

Norwegian Petroleum, 'Troll'. Accessed: 14.06.2023. Available: https://www.norskpetroleum.no/fakta/felt/troll

NPR, ‘This is what the world looks like if we pass the crucial 1.5-degree climate threshold’, Accessed:14.06.2023. Available: https://www.npr.org/2021/11/08/1052198840/1-5-degrees-warming-climate-change

Oni, A. O., Anaya, K., Giwa, T., Di Lullo, G., Kumar, A. (2022) ‘Comparative assessment of blue hydrogen from steam methane reforming, autothermal reforming, and natural gas decomposition technologies for natural gas-producing regions’, *Energy Conversion and Management* (254), 2022, 115245. https://doi.org/10.1016/j.enconman.

Rapier, R, (2020) ‘Life Cycle Emissions of Hydrogen’, *Climate*, Accessed: 14.06.2023. Available: https://4thgeneration.energy/life-cycles-emissions-of-hydrogen/

Reuters, 'Environment'. Accessed: 14.06.2023. Available: https://www.reuters.com/business/environment/norways-climate-choice-old-oil-gas-fields-switch-green-power-or-close-early-2022-08-15/

Schrag, D. P. (2007) ‘Preparing to capture carbon’, *Science*, 315 pp. 812-813.

Statistics, Norway, 2023, 'Utslip til luft’. Accessed: 14.06.2023. Available: https://www.ssb.no/natur-og-miljo/forurensning-og-klima/statistikk/utslipp-til-luft

Szulczewski, M., Juanes, R. (2009) ‘A simple but rigorous model for calculating CO2 storage capacity’, *Energy Procedia*, 1.

United Nation Climate change, 'Climate Plans Remain Insufficient: More Ambitious Action Needed Now'. Accessed: 14.06.2023. Available: https://unfccc.int/news/climate-plans-remain-insufficient-more-ambitious-action-needed-now

The Impact of Autonomous Inflow Control Valve on Enhanced Oil Recovery in SAGD Application

Farhan Hasin Alam [a], Seyed Amin Tahami [a], Nora C.I. Furuvik [a], Britt M.E. Moldestad [a], Soheila Taghavi [a & b, *]

[a] *Department of Process, Energy and Environmental Technology, University of South-Eastern Norway, Norway*
[b] *InflowControl AS, Norway*

*corresponding author: soheila.taghavi.hosnaroudi@gmail.com
soheila.t.hosnaroudi@usn.no

Abstract
The demand for non-conventional oil has increased globally. Non-conventional oil is categorized as extra heavy oil and bitumen. In reservoirs with extra heavy oil and bitumen, thermal methods are used to reduce the oil viscosity. Steam assisted gravity drainage (SAGD) is a thermal recovery method to enhance the bitumen recovery. In this method, steam is injected to bitumen and heavy oil to reduce the viscosity and make the oil mobile and extractable. To obtain an efficient SAGD process, the residence time for steam in the reservoir must be long enough for the steam to condense and release the latent energy to be transferred to the cold bitumen. Early breakthrough of steam in some parts of the well will eventually limit the oil production and must be avoided. Autonomous inflow control valve (AICV) can prevent the steam breakthrough and restrict the excessive production of steam. The objective of this paper is to investigate the performance of AICV and its impacts on increased oil production in a SAGD production well. This is achieved by focusing on the implementation, and performance evaluation of inflow control devices (ICDs) and AICVs compared with standard well perforations. CMG STARS, a multi-phase, multi-component thermal reservoir simulator, is used to perform numerical simulation studies. The simulation results demonstrate the significant benefit of AICV in steam reduction compared to ICD and well perforations. The simulation results demonstrate that utilizing AICV in a SAGD reservoir will lead to higher oil production, less steam production, and a more uniform temperature distribution, and steam chamber conformance. Reduction in steam production, will improve the overall SAGD operation performance. This will also result in more cost-effective oil production, as less steam is needed to be generated for production of each barrel of oil.

1. Introduction

The demand for non-conventional oil has increased globally. Non-conventional oil is categorized as extra heavy oil and bitumen. The mobility of bitumen is quite poor since the viscosity can be as high as 10^6 cP (Ghahfarokhi et al., 2012). In reservoirs with extra heavy oil and bitumen, thermal methods such as steam assisted gravity drainage (SAGD) are used to reduce the oil viscosity and make the bitumen mobile and extractable. More than 80 percent of the world's annual output of heavy oil is accomplished through the utilization of this technique (Xu et al., 2020).

The SAGD well configuration typically consists of a pair of horizontally aligned wells that are between 500 and 1000 meters in length (Shen, 2013). The top wellbore is utilized for steam injection which is located about 4-6m above the production well, and the lower wellbore is utilized for oil production. When horizontal wellbores are used in SAGD, reservoir contact, and the overall well productivity are both significantly improved (Shen, 2013). Steam is injected into the reservoir from the injection well. This creates steam chambers at the interfaces as shown in Figure 1. These steam chambers expand both vertically and laterally (Shen, 2013). Latent heat from the steam is transferred to the bitumen at the interface, making it less viscous and more mobile. Due to the action of gravitational forces, the steam condensate, and the mobile bitumen flow downwards into the producer well. SAGD has been shown to be a successful and cost-effective way to get bitumen out of heavy oil reservoirs.

Several technologies have been created to optimize the performance of the SAGD process. Advanced well completion devices such as inflow control devices (ICDs) and autonomous inflow control valves (AICVs) can be used for improving bitumen recovery in SAGD operations. ICDs are intended to control fluid flow within a wellbore and assuring uniform distribution of steam. AICVs, on the other hand, are adjustable valves that regulate the openings automatically based on the fluid viscosities. The AICVs are thereby preserving

balanced production rates and are controlling the inflow profiles.

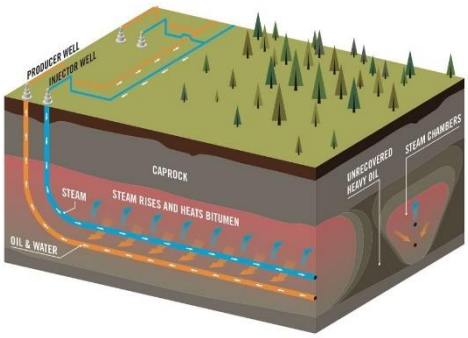


Figure 1: Principles of SAGD operation (staff, 2016)

The use of ICDs and AICVs in SAGD operations has the potential to solve several issues associated with steam injection and bitumen recovery. These issues include steam chamber conformance, early steam breakthrough, irregular fluid distribution, water and gas coning, and excessive production of undesirable fluids. By strategically employing ICDs and AICVs, operators can optimize thermal performance, maximize bitumen recovery, and reduce operating expenses.

The objective of this paper is to investigate the performance of AICVs, and its impacts on increased oil production in a SAGD production well. This is achieved by focusing on the implementation, and performance evaluation of ICDs and AICVs compared with standard well perforations. The novelty of this work is to simulation of the AICV and ICD behavior in a dynamic reservoir simulator under SAGD conditions. The functionality of the AICV and ICD is simulated through tabulated data based on the experiments presented in previous author's work (Taghavi et al., 2022). CMG STARS, a multi-phase, multi-component thermal reservoir simulator, is used to perform numerical simulation studies.

2. Advanced Wells with Inflow Control Technologies; ICD and AICV

Inflow control technologies such as ICDs and AICVs were introduced to the oil industry to overcome the early water and gas breakthrough challenges associated with the heel-toe effect in horizontal wells. Drilling long horizontal wells can increase reservoir contact, resulting in improved oil recovery. The pressure difference between the toe and heel sections of the well becomes large in long horizontal wells due to the pressure drop induced by friction between the inner pipe surface and fluid flowing through the pipe. This pressure difference in the well generates a higher pressure drawdown between the wellbore and the reservoir at the heel than at the toe, resulting in a greater inflow of reservoir fluid in the heel rather than in other areas of the well as shown in Figure 2. This phenomenon is known as the heel-toe effect. Because of heel-toe effect, early breakthrough of water and/or gas occurs at the heel section of the well, decreasing oil recovery efficiency.

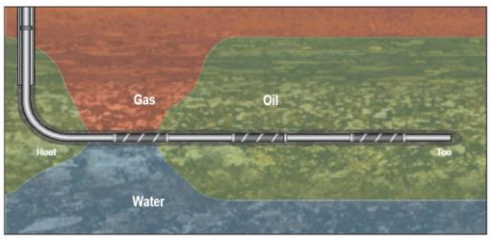


Figure 2: Gas and water breakthrough at the heel section of the well (Ellis et al., 2009).

2.1. ICDs in advanced wells

ICD is used to restrict the flow of fluid entering the base pipe from the annulus. It is a passive inflow control device, meaning it has no active components that can be regulated or altered to regulate the flow through it.

The governing equation of the nozzle-type ICD, as shown in Figure 3, is as follows (Lauritzen & Martiniussen, 2011):

$$\Delta P = \frac{8\rho Q^2}{d^4\pi^2 n^2 {C_D}^2} \qquad (1)$$

Where ΔP is the pressure drop through the nozzle, ρ is the fluid density, Q is the volumetric flow rate of the fluid through the nozzle, d is the diameter of the nozzle, n is the number of tested nozzles, and C_D is the discharge coefficient. C_D is mostly a function of the Reynolds number (Re) (Lauritzen & Martiniussen, 2011).

The pressure drop through the nozzle is mainly dependent on the fluid density.

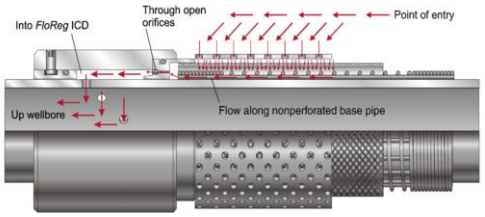


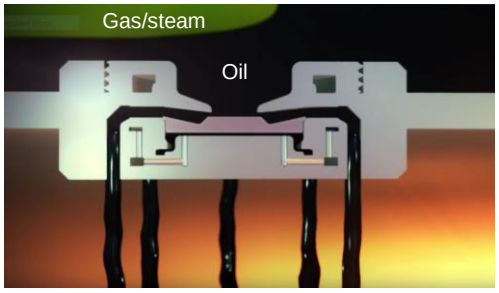
Figure 3 : Nozzle-type ICD (Birchenko et al., 2010).

2.2. AICVs in advanced wells

AICV is a novel inflow control system that combines the most advantageous characteristics of inflow controllers. AICVs are autonomous, meaning that they operate without the need of external control systems and constant human involvement. For oil production, AICV offers minimal flow restriction and the capability to close for water and gas/steam while simultaneously producing oil from other zones along the well. The valves in zones where gas/steam

and water break through into the well, will close locally. Figure 4 shows AICV in closed and open position. Figure 4a shows that valve is open and producing oil as gas/steam is approaching the valve. Figure 4b illustrates that the gas/steam has reached the valve inlet, and valve is closed for gas/steam.

(a)



(b)

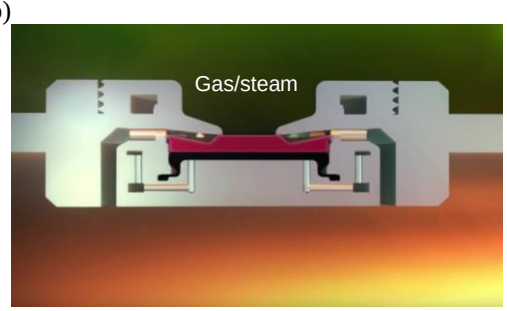


Figure 4: AICV in open (a) and closed (b) position (Aakre et al., 2018).

The mathematical model describing the performance of the AICV can be described as:

$$\Delta P_{Tot} = \left(\frac{\rho_{mix}^2}{\rho_{cal}}\right) \cdot \left(\frac{\mu_{cal}}{\mu_{mix}}\right)^{y} \cdot a_{AICD} \cdot Q^{x} \qquad (2)$$

where ΔP_{Tot} is the differential pressure across the AICV, ρ_{cal} and μ_{cal} are the calibration fluid density and viscosity, and ρ_{mix} and μ_{mix} are the mixture fluid density and viscosity respectively. The parameter a_{AICD} is a valve characteristic given by the ICD strength, Q is the volumetric mixture flow rate, and x and y are constants (Taghavi & Ghaderi, 2021). It can be interpreted from equation (2) that the pressure drop through the AICV is much more viscosity dependent than density dependent. The concept and principle of AICV is described in detail in earlier scientific works (Aakre, 2017; Aakre et al., 2013).

3. Reservoir and Wellbore Model in CMG

CMG 2022.10 general release by Computer Modeling Group Ltd. is used for accomplishing the objectives of this paper. The software has thirteen modules, each for a specific purpose. Reservoir grid modeling, well modeling, creation of fluid models, rock fluid properties and importing previously created well, reservoir and component properties are done using the Builder module of the CMG software. CMG STARS is responsible for conducting thermal and steam additive simulations. Thermal oil recovery methods such as SAGD, can be simulated with the help of STARS.

3.1. Reservoir construction in CMG Builder

A cuboid shaped reservoir has been considered where gravitational force is acting along the k-direction (vertical). The reservoir grid building constraints have been shown in Table 1.

Table 1:The dimensions of the drainage area.

Direction	No. of Blocks	Block size distributions (No. of blocks*block length) [m]
I (x)	30	30*50
J (y)	15	2*20, 10, 8, 5, 4, 3, 1, 3, 4, 5, 8, 10, 2*20
K (z)	20	11*3.5, 1, 2*3, 1, 3*3, 2*5

The areas marked in blue in Figure 5 represent the location of the injector and producer wells. All the cells within the blue area are very small in dimensions.

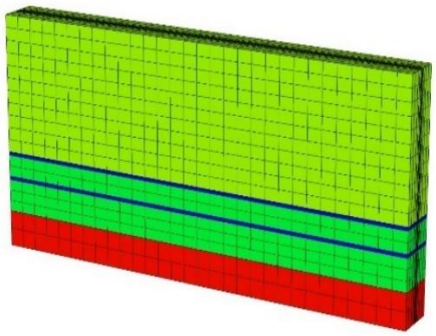


Figure 5: Reservoir 3D view

The details regarding the reservoir characteristics and parameters for initialization are presented in Table 2.

Table 2: Reservoir Characteristics initialization details.

Property	Value
Porosity	30 %
Rock wettability	Water wet
Reservoir top depth	400 m
Initial pressure at top of the reservoir	1500 kPa
Initial temperature	12°C
Initial water saturation	0.10
Reference depth	430 m
Water-oil contact depth	455.5 m
Oil mole fraction (dead oil)	0.80
Oil mole fraction (solution gas)	0.20

3.1.1. Reservoir rock and fluid properties

The reservoir rock and fluid thermal properties are given in Table 3.

Table 3: Reservoir rock and fluid thermal properties.

Property	Value
Formation compressibility	2.90×10^{-6} 1/kPa
Rock volumetric heat capacity	2.35×10^{6} J/(m^3·C)
Rock thermal conductivity	6.60×10^{5} J/(m·day.C)
Oil thermal conductivity	1.25×10^{4} J/(m.day.C)
Water thermal conductivity	5.35×10^{4} J/(m.day.C)
Gas thermal conductivity	3.20×10^{3} J/(m.day.C)
Over/Under-burden volumetric heat capacity	2.35×10^{6} J/ (m^3.C)
Over/Under-burden thermal conductivity	1.496×10^{5} J/(m.day.C)

The viscosity changes of bitumen as a function of temperature is taken from the experimental work of (Ghahfarokhi et al., 2012) and is shown in Figure 6. The oil viscosity at standard pressure decreases radically with the increase in temperature.

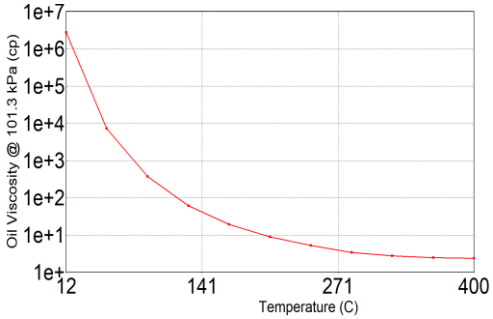


Figure 6:Viscosity of Athabasca bitumen sample versus temperature.

Generally, it is challenging to obtain information about relative permeability for different fields. Data for relative permeabilities are set manually in table form in CMG Builder. Two-phase relative permeabilities for liquid-gas and water-oil are shown in Figure 7 and Figure 8 respectively. The datasets have been calculated based on the Stone II model.

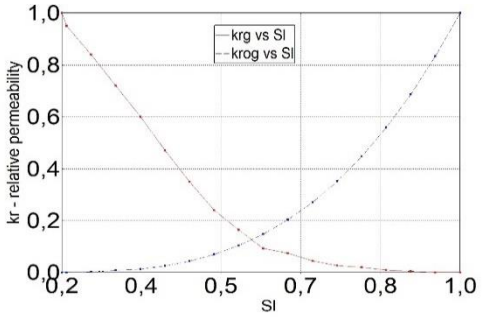


Figure 7: Liquid - gas relative permeability curves.

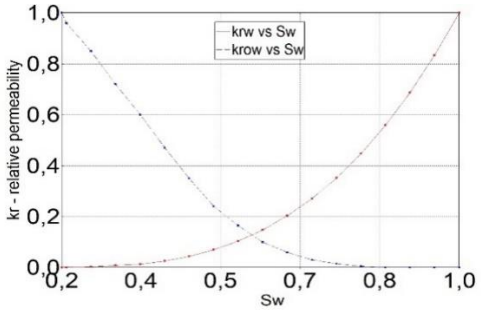


Figure 8: Water-oil relative permeability curves.

Sl is the liquid saturation, Kr is the endpoint relative permeability, krg is the relative permeability to gas at Sl, krog is the relative permeability to oil in the presence of gas at liquid saturation Sl = 1-Sg. In addition, Sw is the water saturation, krw is the relative permeability to water at Sw and krow is the Relative permeability to oil at Sw.

3.2. Well modelling in Builder

The simulation time has been set to 10 years for the SAGD operation and these 10 years have been divided into two phases. The first phase, also known as the circulation period, starts from 1st of January 2023 and continue for six months until 1st of July 2023. The SAGD period starts from 1st of July 2023 and continue until 1st of January 2033. Each well is 1201 meters long horizontally. Eight wells have been defined for accomplishing the whole SAGD process. Their names and period and mode of operation are shown in Figure 9. The FlexWell (FW) model is developed by CMG and is used to model the fluid flow in the wellbore and between the wellbore and the reservoir. FW is an advanced discretized mechanistic wellbore model which models the complex well completions (Mohd Ismail et al., 2021). The injector FWs are placed 6 meters above the producer FWs maintaining the optimum vertical distance. Figure 10 and Figure 11present the wells trajectories during the circulation phase and the SAGD phase respectively.

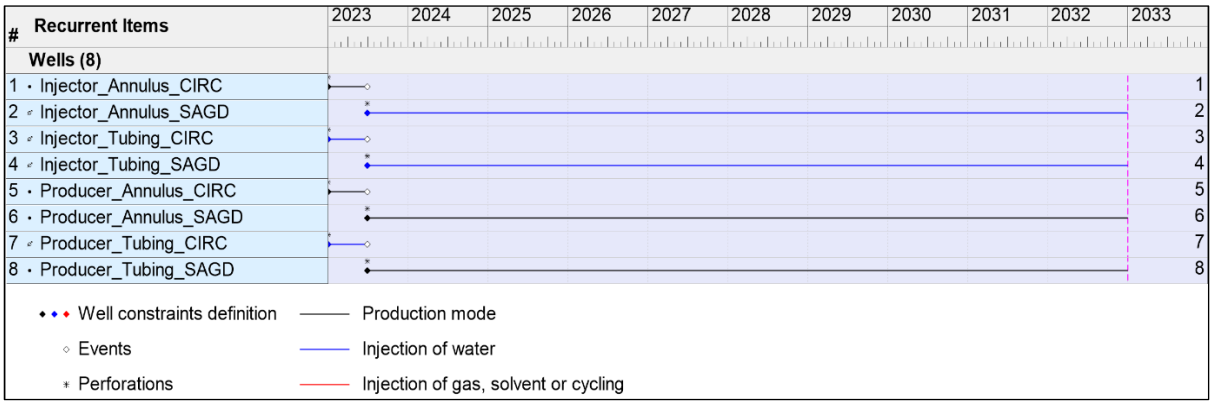


Figure 9: Timeline view of well operation data.

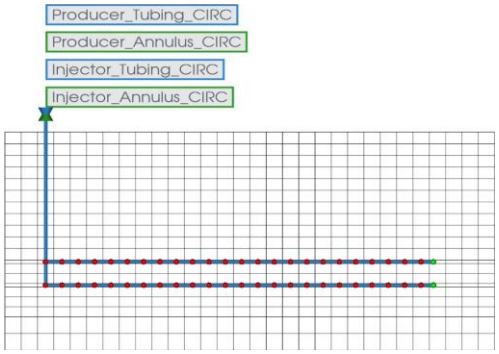


Figure 10: Well trajectories during circulation phase.

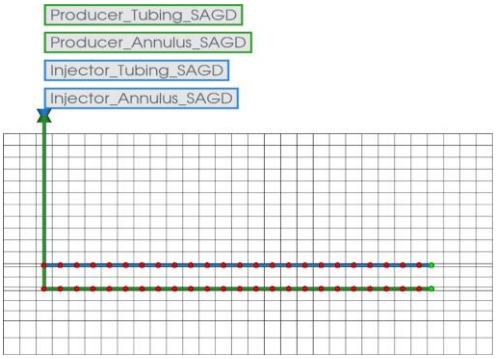


Figure 11:Well trajectories during SAGD phase.

The wells have been modeled this way in order to maintain similarity with the real-world conditions as the wells working during the circulation stage are completely turned off as the SAGD stage starts, rendering them nonexistent by CMG. Essentially, the wells present during the circulation stage are neither active nor present during the SAGD stage.

The well constraints for circulation and SAGD phases are listed in Table 4 and Table 5.

Table 4 : Circulation period well constraints.

FWs	Name	Function	Constraints
Injector FW_ CIRC	Injector_ Annulus _CIRC	Producer	MIN BHP 3500 kPa
			MAX STL 150 m^3/day
	Injector_ Tubing _CIRC	Injector	MAX BHP 4000 kPa
			MAX STW 100 m^3/day
			Injection temperature 250°C
			Steam quality 0.9
Producer FW_ CIRC	Producer_ Annulus _CIRC	Producer	MIN BHP 3500 kPa
			MAX STL 150 m^3/day
	Producer_ Tubing_ CIRC	Injector	MAX BHP 4000 kPa
			MAX STW 100 m^3/day
			Injection temperature 250°C
			Steam quality 0.9

Table 5: SAGD period well constraints.

FWs	Name	Function	Constraints
Injector FW_ SAGD	Injector_ Annulus_ SAGD	Injector	MAX BHP 4000 kPa
			MAX STW 500 m^3/day
			Injection temperature 250°C
			Steam quality 0.9
	Injector_ Tubing_ SAGD	Injector	MAX BHP 4500 kPa
			MAX STW 500 m^3/day
			Injection temperature 250°C
			Steam quality 0.9
Producer FW_ SAGD	Producer_ Annulus_ SAGD	Producer	MIN BHP 2000 kPa
			MAX STL 1500 m^3/day
	Producer_ Tubing_ SAGD	Producer	MIN BHP 2000 kPa
			MAX STL 1500 m^3/day

BHP is bottom hole pressure, STW is the surface water rate, and STL is the surface liquid rate.
The dimensions of the annulus and tubing are listed in Table 6.

Table 6: Diameters of annulus and tubing.

Parameter	Size
Tubing wall inner diameter	0.104 m
Tubing wall outer diameter	0.114 m
Annulus wall inner diameter	0.166 m
Annulus wall outer diameter	0.177 m

4. Results and Discussions

The main well of interest for this study is the Producer_Annulus_SAGD, the annulus of Producer FW_SAGD. Depending on the case definitions, Producer_Annulus_SAGD annulus will either have only perforations or be equipped with ICDs or AICVs for comparing oil recovery. The rest of the wells will operate with well perforations.

4.1 Simulation cases

Six cases have been established for simulation purposes. The simulation cases are well perforations (without any inflow controllers), well completed with 4 ICDs in each 50 meters of the horizontal length, and well completed with 4 AICVs in each 50 meters of the horizontal length in both homogenous and heterogeneous reservoir. The horizontal permeability of the homogeneous reservoir is 1800 mD in all blocks. The permeability distribution of the heterogeneous reservoir is illustrated in Figure 12.

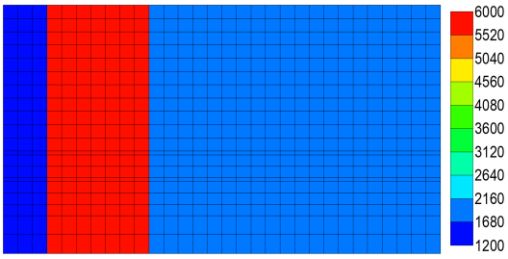


Figure 12: Permeability distribution of heterogeneous reservoir (I-K plane view).

4.2 Simulation results in the homogenous reservoir

During the circulation period, steam is injected from both wells. This is to establish thermal communication between the injector and producer and warm up the reservoir. Figure 13 shows the temperature distribution at the end of the circulation period which is between 70-100 ºC.

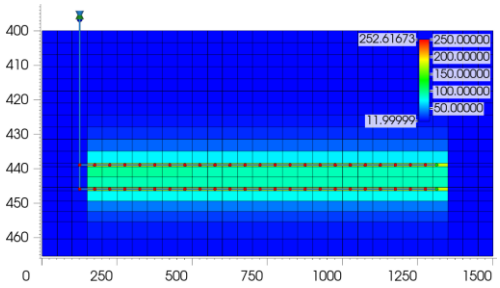


Figure 13: Temperature distribution at the end of the circulation period (I-K plane view).

In order to study the performance of ICD and AICV, the accumulated oil, gas (steam), and water for the AICV and ICD completion cases are compared to the case without any inflow controllers (perforations), see Figure 14. Under homogeneous conditions, the perforation case (red line) falls behind the case with ICDs (dashed green line) and the case with AICVs (solid green line), having the lowest cumulative oil production for the highest cumulative gas and water production. Based on cumulative oil and gas production, the AICV case outperforms the ICD case having higher oil production and lower gas production as shown in Figure 14.

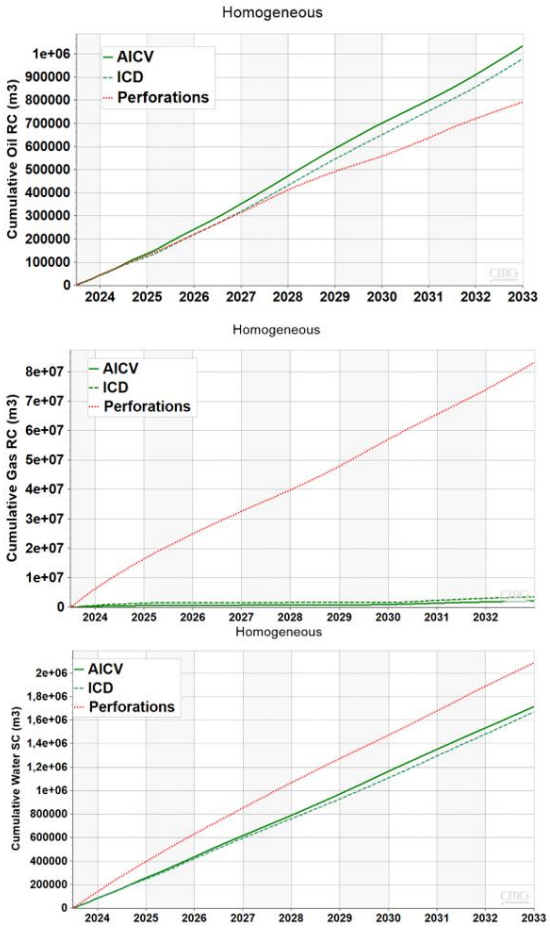


Figure 14 : Cumulative oil, gas, and water production for perforations, ICDs and AICVs in a homogenous reservoir.

Understanding the formation of steam chambers and temperature distribution across the reservoir is an important aspect of the SAGD process. Steam chamber patterns and temperature distributions can also be used to indicate in which case there is more steam production. From Figure 15, it can be interpreted that due to the high steam production for well perforations, the steam chamber has not reached the maximum temperature after 5 years, see Figure 15a. Looking closely along J-K plane shown in Figure 15c, illustrates that the AICV case has a slightly better steam distribution than the ICD case shown in Figure 15b. This means that AICVs are better in handling steam breakthroughs than ICDs and well perforations.

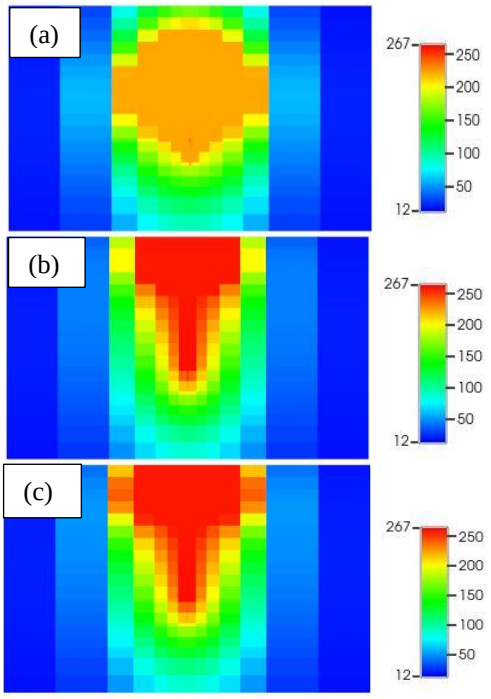


Figure 15 : Steam chamber conformance along J-K plane for (a) perforations, (b) ICDs, and (c) AICVs.

4.3 Simulation results in the heterogeneous reservoir

Figure 16 illustrates that the AICV case outperforms the ICD and perforations cases in terms of having the highest cumulative oil production and the least cumulative gas and water production. The perforation case (red line) falls behind the case with ICDs (dashed green line) and the case with AICVs (solid green line), having the lowest cumulative oil production for the highest cumulative gas and water production.

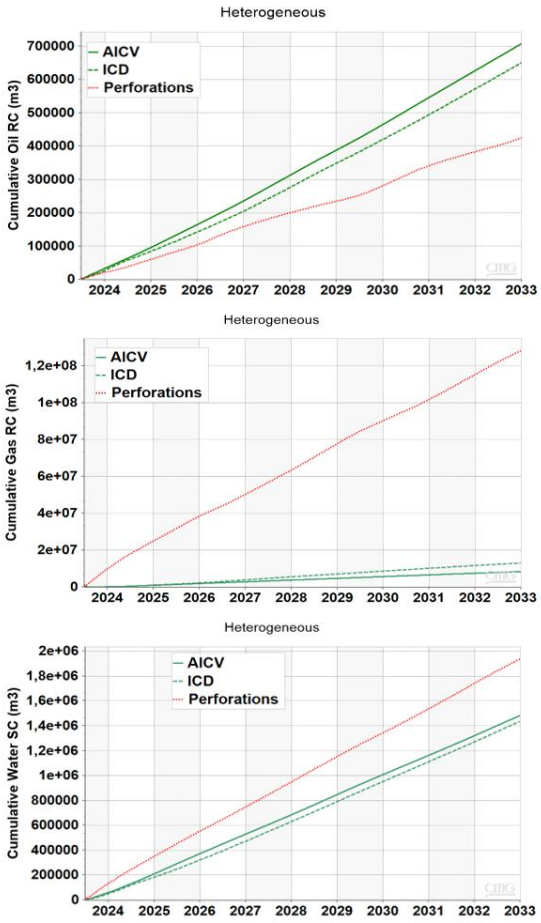
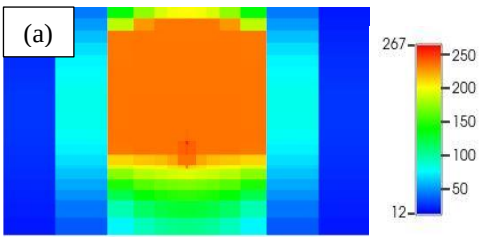


Figure 16: Cumulative oil, gas, and water production for perforations, ICDs and AICVs in a heterogeneous reservoir.

This directly indicates that AICVs are better in recovery of heavy oil and in resisting gas (steam), and water production compared to ICDs and well perforations when subjected to heterogeneous reservoir conditions, similar to that under homogenous conditions.

As illustrated by Figure 17a, the perforation case does not have a uniform steam chamber conformance and temperature distribution after 5 years. Both ICDs and AICVs show uniform steam chamber conformance which has reached maximum temperature after 5 years as shown in Figure 17b and Figure 17c respectively.



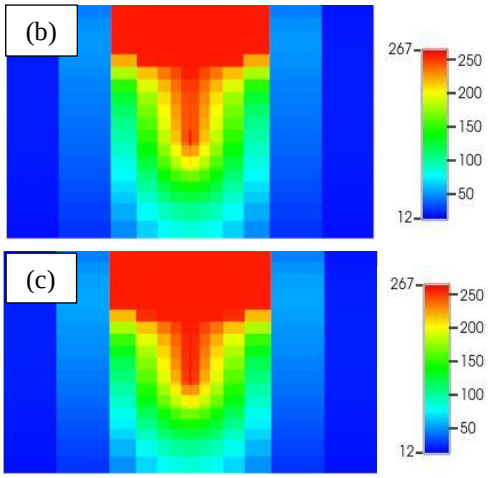


Figure 17 : Steam chamber conformance along J-K plane for (a) perforations, (b) ICDs, and (c) AICVs.

Analysis of Figure 15 and Figure 17 demonstrates that ICDs and AICVs are much better in maintaining proper temperature distribution across the reservoir and in formation of uniform steam chamber compared to well perforations. When looked from the J-K plane, it is seen that steam chamber conformations with well perforations in both homogeneous and heterogeneous cases do not reach the maximum steam injection temperature. On the contrary, ICDs and AICVs evidently show better steam chamber conformance by reaching maximum injection temperatures.

5. Conclusions

The impact of AICV on enhanced oil recovery in a SAGD production well is investigated. This is achieved by developing a wellbore-reservoir model in the CMG STARS simulator.

Both homogenous and heterogeneous reservoirs are studied by considering ICD and AICV completion and well perforations only.

The simulation results demonstrate that utilizing AICV in a SAGD reservoir will lead to higher oil production, less steam production, and a more uniform temperature distribution, and steam chamber conformance.
Reduction in steam production, will improve the overall SAGD operation performance. This will also result in more cost-effective oil production, as less steam is needed to be generated for production of each barrel of oil. Less steam generation means less energy demand, that consequently contribute to lower the intensity of greenhouse gas emissions.

References

Aakre, H. (2017). *The impact of autonomous inflow control valve on increased oil production and recovery* University College of Southeast Norway, Faculty of Technology, Natural Sciences and Maritime Sciences]. Kongsberg.

Aakre, H., Halvorsen, B., Werswick, B., & Mathiesen, V. (2013). Smart Well With Autonomous Inflow Control Valve Technology. SPE Middle East Oil and Gas Show and Conference, https://doi.org/10.2118/164348-MS

Aakre, H., Mathiesen, V., & Moldestad, B. (2018). Performance of CO2 flooding in a heterogeneous oil reservoir using autonomous inflow control. *Journal of Petroleum Science and Engineering*, *167*, 654-663. https://doi.org/https://doi.org/10.1016/j.petrol.2018.04.008

Birchenko, V. M., Muradov, K. M., & Davies, D. R. (2010). Reduction of the horizontal well's heel–toe effect with inflow control devices. *Journal of Petroleum Science and Engineering*, *75*(1), 244-250. https://doi.org/https://doi.org/10.1016/j.petrol.2010.11.013

Ellis, T., Erkal, A., Goh, G., Jokela, T., Kvernstuen, S., Leung, E., Moen, T., Porturas, F., Skillingstad, T., Vorkinn, P. B., & Raffn, A. G. (2009). Inflow control devices - Raising profiles. *21*, 30-37. https://doi.org/https://studylib.net/doc/18336329/inflow-control-devices%E2%80%94raising-profiles

Ghahfarokhi, A. J., Jelmert, T. A., Kleppe, J., Ashrafi, M., Souraki, Y., & Torsaeter, O. (2012). Investigation of the Applicability of Thermal Well Test Analysis in Steam Injection Wells for Athabasca Heavy Oil. SPE Europec/EAGE Annual Conference, https://doi.org/10.2118/154182-MS

Lauritzen, J. E., & Martiniussen, I. B. (2011). Single and Multi-phase Flow Loop Testing Results for Industry Standard Inflow Control Devices. SPE Offshore Europe Oil and Gas Conference and Exhibition, https://doi.org/10.2118/146347-MS

Mohd Ismail, I., Mathiesen, V., Abraham, A., Ranjbar, E., Zarei, F., & Walter, J. (2021). An Innovative Modelling Approach in Characterization of Autonomous Inflow Control Valve Performance to Maximizing Oil Recovery in Heavy Oil-SAGD Application. SPE Thermal Well Integrity and Design Symposium, https://doi.org/10.2118/203859-MS

Shen, C. (2013). Chapter 17 - SAGD for Heavy Oil Recovery. In J. J. Sheng (Ed.), *Enhanced Oil Recovery Field Case Studies* (pp. 413-445). Gulf Professional Publishing. https://doi.org/https://doi.org/10.1016/B978-0-12-386545-8.00017-8

staff, J. (2016, 2016). *Temporarily shutting in oilsands SAGD wells might not be as harmful as you think*. Retrieved February from https://www.jwnenergy.com/article/2016/5/26/temporarily-shutting-oilsands-sagd-wells-might-not/

Taghavi , S., & Ghaderi, A. (2021). On Uncertainty Analysis of the Rate Controlled Production (RCP) Model. First SIMS EUROSIM Conference on Modelling and Simulation, Finland. https://doi.org/10.3384/ecp21185271

Taghavi, S., Aakre, H., & Moldestad, B. M. E. (2022). Performance Analysis of Autonomous Inflow Control Valve in a SAGD Late Life Process with Non-Condensable Gases. SPE Canadian Energy Technology Conference, https://doi.org/10.2118/208915-MS

Xu, Z.-X., Li, S.-Y., Li, B.-F., Chen, D.-Q., Liu, Z.-Y., & Li, Z.-M. (2020). A review of development methods and EOR technologies for carbonate reservoirs. *Petroleum Science*, *17*(4), 990-1013. https://doi.org/10.1007/s12182-020-00467-5

Estimation of effluent nutrients in municipal MBBR process

Tiina M. Komulainen[a,*] Abdul Malik Baqeri[a] Einar Nermo[a] Arvind Keprate[a] Torgeir Saltnes[b] Katrine Marsten Jansen[c] Olga Korostynska[a]

[a]*Oslo Metropolitan University, Norway,* [b]*Hias How2O, Norway,* [c]*Hias IKS, Norway*
[*]tiina.komulainen@oslomet.no

Abstract

The recently updated European Union's Urban Waste Water Treatment Directive proposal, European Green Deal, Biodiversity Strategy for 2030, and EU's Energy System Integration highlight a pressing need for innovative biological nutrient removal processes and energy-efficient control methods to reduce pollution and minimize the carbon footprint at water resource recovery facilities. The aim of the PACBAL research project is to develop estimation methods for nutrient profile in a novel industrial Moving Bed Biofilm Reactor (MBBR) process. This study devises and assesses a range of data-driven methods to estimate effluent phosphorus concentration by utilizing a combination of real sensors with software models. The resulting virtual sensor could facilitate the design of energy-efficient control strategies. The case study data are collected from the MBBR process at Hias water resource recovery facility in Norway. Data sets from December 2022 to March 2023 include varying weather conditions, such as rain, dry, and snow. The Hias Process consists of three anaerobic and seven aerobic zones, where biomass carriers removes over 90 percent of the phosphorus from the wastewater in simultaneous biological processes. The industrial online measurements include wastewater flowrate, aeration rates, dissolved oxygen and nutrients COD and NO_2/ NO_3 at inlet and total suspended solids at outlet. Dynamic data-driven models indluding transfer functions, state-space models and ARX models, were developed and compared to estimate the outlet phosphorus concentration. Model fitness to validation data was around 7% with ARX models, and up to 18% with tranfer function models and state-space models. The first and second order models gave similar results. The state-space models will be developed further and implemented to into virtual sensors that will enable energy-efficient control strategy development.

1 Introduction

There is a significant demand for novel biological nutrient removal processes and energy-effective control methods for minimization of carbon footprint and environmental pollution at wastewater resource recovery facilities (WRRF). As the European Commission has proposed an updated urban waste water treatment directive (EuropeanCommission, 2022), stricter requirements will be set for the removal of nutrients such as phosphorus, carbon and nitrogen. Water resource recovery facilities use approximately one percent of the total energy consumption in the European union. EU plan on energy system integration (EuropeanCommission, 2021) requires actions on energy efficiency that are necessary to convert the WRRFs from an energy consumer to energy-neutral user, or even an energy producer. At the municipal WRRF the primary wastewater treatment, clarification, is followed by a secondary treatment process, which removes nutrients such as phosphorus, carbon and nitrogen. The secondary treatment process relies either on chemical additions or biological process. Hias How2O has developed a novel continuous-flow moving bed biofilm reactor process with enhanced biological phosphorus removal (MBBR-EBPR) and simultaneous nitrification and denitrification as described in Rudi et al. (2019). This process is an alternative to the biological process with an activated sludge. In the Hias Process, large amounts of small biofilm carriers, submerged in the wastewater, circulate through the ten process stages while the nutrients are removed from the wastewater in simultaneous biological processes in different layers of the biofilm. Soluble phosphorus (PO_4) in wastewater is removed biologically by phosphorus accumulating organisms (PAO) that grow on biofilm carriers.

The first step towards energy-effective control is development of dynamic models between the Hias Process inputs and effluent nutrient composition. Hence, we need to develop a dynamic model that can reproduce effluent nutrient composition accurately enough based on the Hias process inputs.

Moving bed bioreactor process can be modeled with ASM2D model as described by Henze et al. (1999). Application of the ASM2D model for the Hias Process with ten stages, would require estimation of 430 parameters and simultaneous solution of 210 rate equations. Hence, ASM2D model in its original form is too complicated with the current instrumentation.

Nair et al. (2019) have simplified the ASM2D model to include 8 components and 7 process rate equations, 2 stoichiometric constants, 11 saturation coefficients and 7 rate constants for both anaerobic and aerobic stages. Therefore, the ASM2D model, as it currently stands, is overly complex given the available instrumentation. The ASM2D model is further simplified for anaerobic basins with 3 components and 3 rate equations in Nair et al. (2020). Nair et al have demonstrated soft-sensor concept with Kalman filter for laboratory pilot system with one tank. Due to the available instrumentation in the Hias Process, this approach cannot be tested either.

Data-driven methods do not require a specific set of instrumentation, and several recent papers have reported successful applications to effluent prediction using multivariate linear regression (Tomperi & Leiviskä, 2018), feedforward-backpropagation networks (El-Rawy et al., 2021), and time-series models and machine learning models (Ly et al., 2022). In a recent master thesis (Nermo, 2023) transfer function models were developed to predict the Hias Process effluent phosphorus. In this article we continue Nermo's work by refining the input variable set and developing time-series models and state-space models.

1.1 Aim and research questions

The main obstacles for development of novel energy and emission efficient control strategies are the scarcity and cost of online measurements of nutrients in municipal WRRFs. Control strategy development requires models that sufficiently reproduce the nutrient dynamics of the plant based on available measurements. Hence, the aim of this work is to develop and compare different data-driven modeling approaches that enable control strategy development. Research question: Which data-driven models are most effective in predicting nutrient variations in the Hias Process effluent?

2 Materials and methods

2.1 Software

Matlab software package version R2023a was used for the simulations and System Identification Toolbox for modeling. The simulation method was ode23s with automatic settings for the time step and error tolerance.

2.2 The Hias Process and instrumentation

The Hias Process with instrumentation is illustrated in Figure 1. The clarified influent and the recirculated biofilm carriers on a conveyor belt enter the anaerobic basins. The water and biofilm carriers float through the process with gravity. The three anaerobic basins

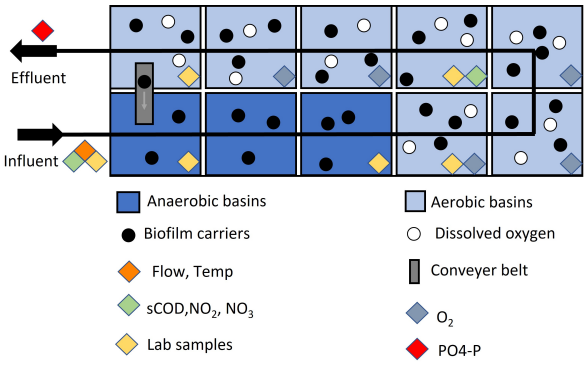


Figure 1. The Hias Process with instrumentation

are mixed to ensure sufficient distribution of biofilm carriers in the water. Aeration in the following seven basins ensures sufficient dissolved oxygen concentrations for aerobic nutrient removal. The Hias Process instrumentation includes continuous measurements of flowrates and nutrients compositions. Influent COD, NO_2 and NO_3 are measured continuously with the spectrophotometric instrument. Effluent phosphorus is measured using an online-analyzer with 10 minutes sampling time. These measurement as listed in Table 1.

Table 1. Online measurements

Symbol	Description	Unit
F	Water flowrate inlet	m^3/h
T	Temperature inlet	^oC
COD	COD inlet, basin7	g/m^3
NOX	$NO_2 + NO_3$ inlet, basin7	g/m^3
FO_i	Aeration rate basin 4-10	m^3/h
DO_i	Oxygen basin 4-10	m^3/h
TSS_{out}	Total suspended solids out Hias Process	g/m^3
TSS_{disc}	Total suspended solids after disc filter	g/m^3
PO_{4out}	PO_4 effluent	g/m^3

2.3 Data collection and pre-processing

The Industrial IoT platform KYB, developed by Digitread Connect, was used for uploading and standardizing operational data. The online data sets and laboratory data sets were collected in .csv format. The outliers in the online data set were first removed. Then, the missing values in the online data sets were filled. The correlations between variables were studied with a heatmap.

2.4 Data-driven dynamic models

The data-driven models suitable for control strategy design are desirable for the project, hence transfer function models, state-space models and time-series models were chosen. These linear models can be developed by following the system identification procedure by (Ljung, 1999). First, the data sets for modeling and validation are chosen. Then, a set of input and output variables is composed, and time delays are calculated between each input variables and output variable. The data is imported to the Matlab System Identification toolbox and model type is chosen. Lower order models are preferable to avoid modeling of noise. The identified models are compared using the fitness index and mean square error.
A first order transfer function model between inputs $U_i(s)$ and output $Y(s)$ consist of gain (K_p), delay (T_d), and time constant (T_{p1}) is presented in Equation 1:

$$TF(s) = \frac{Y(s)}{U_i(s)} = \frac{K_p}{(T_{p1}s+1)}e^{-T_d s} \tag{1}$$

The time series model, an auto-regressive model with exogenous inputs (ARX) are presented in the equation 2. The model output $y(t)$ has order of 2, and the model inputs $u(t)$ have order of 2 with time delay td.

$$\begin{aligned} y(t) - a_1 y(t-1) - a_2 y(t-2) \\ = b_{i1}u(t-td) + b_{i2}u(t-td-1) + e(t) \end{aligned} \tag{2}$$

A state-space model is presented in equation 3. The derivative of the state x(t) is related to the model inputs u(t) with delays td via coefficient matrices A and B, and to the model error e(t) with coefficient K. The measurement y(t) is related to the state x(t) via matrix C. The error e(t) is calculated as difference between the model prediction and the real measurement. The measurement y(t) is not affected by inputs u(t) and hence, D matrix is zero.

$$\begin{aligned} \frac{dx(t)}{dt} &= Ax(t) + Bu(t-td) + Ke(t) \\ y(t) &= Cx(t) + Du(t-td) + e(t) \end{aligned} \tag{3}$$

The data-driven models are compared with each other using the fitness index (FIT) between the real measurements $y_{i,meas}$, the mean values of the real measurements $y_{i,mean}$, and the model calculated output $y_{i,model}$ in Equation 4, where norm is the Euclidean norm.

$$FIT = (1 - \frac{norm(y_{i,meas} - y_{i,model})}{norm(y_{i,meas} - y_{i,mean})})100 \tag{4}$$

Table 2. Pre-processing of input and output variables

Variable	Scaling	Scaled mean	Scaled stdev
F	$3.6/1000$	0.3184	0.0777
COD	$/1000$	0.4602	0.1015
NOX	$/10$	0.3566	0.1548
$FO5$	$/1000$	1.7614	0.7151
$FO8$	$/1000$	0.7833	0.2815
ΔTSS	$/1000$	0.1552	0.0323
PO_{4out}	no scaling	0.2236	0.1024

3 Results

3.1 Data collection and selection of test data

The data was screened and quality of measurements were assessed. Due to many missing measurements, inlet temperature, COD and NOX in basin 7 were omitted from the data set. The estimation period was 1.12.2022-31.1.2023 and the validation period was 1.2.2023-31.3.2023.

3.2 Data pre-processing

The outliers in the data set were identified based on the three standard deviation rule, and removed. Then, the missing values in the data sets were filled in using Matlab knnimpute function based on nearest-neighbor imputation method. The data was analyzed with pair plot (not included here), which shows that the data is highly nonlinear, and thus z-normalization cannot be applied. The data were scaled by scalar multiplication as given in Table 2 and the means (Table 2) were removed.

3.3 Output variable

The only available output variable in the data sets is effluent phosphorus concentration PO_{4out}.

3.4 Input variable selection

The measured input variables include: inlet wastewater flowrate F, inlet carbon COD, inlet nitrate and nitrite NOX, aeration rates FO_4 ... FO_{10}.The dissolved oxygen concentration in aerobic basins DO_4... dO_9 are dependent on aeration rate. Results of Nermo's Master's Thesis (Nermo, 2023) pointed into careful selection of input variables, and creation of extra variable representing the biomass storage capacity. The biomass carrier's storage capacity is based on phosphorus removal during previous biomass carrier cycle around the ten basins. The storage capacity was modeled with the difference of total suspended solids between treated water out of the Hias Process TSS_{out} and

the total suspended solids after the disc filter TSS_{disc}, as in Equation 5:

$$\Delta TSS(t) = TSS_{out}(t) - TSS_{disc}(t) \quad (5)$$

A heat map of correlations between the variables was plotted in Figure 2.
The heat map shows that aeration rates in basins 4 - 10 (FO_4... FO_{10}) are highly correlated with each other and PO_{4out}. Hence, we selected two aeration rates, $FO5$ and $FO8$ for input variables representing the manipulated variables of the system. Dissolved oxygen measurements are not included as these are state variables and highly correlated with the aeration rates. Inlet wastewater flow rate (F), inlet carbon (COD), inlet nitrite and nitrate concentration (NOX) and total suspended solids (ΔTSS) are the measured disturbance variables of the system. The biological mechanism of (COD) is improving phosphorus removal whereas (NOX) would hinder phosphorus removal. It is surprising that both are correlated positively with PO_{4out}, hence both are included as input variables. Storage capacity (ΔTSS) is mildly correlated with PO_{4out}, and included as input variable. The inlet flowrate (F) should have an effect on the PO_{4out}, but the correlation is low. However, (F) is included as input variable. Based on the biological phosphorus removal process (MBBR-EBPR), we could expect that increased flowrate would dilute PO_{4out} (- sign), NOX to hinder removal and increase the PO_{4out} (+ sign), aeration rates (FO_i to improve the removal and decrease the PO_{4out} (- sign), and storage capacity ΔTSS to improve the removal and decrease the PO_{4out} (- sign). Based on laboratory measurements, inlet COD is very correlated with inlet phosphorus PO_{4in}, and increase in COD means increase in PO_{4in}, which in turn increases PO_{4out} (+sign). However, the heat map in Figure 2 gives positive correlations between PO_{4out} and the selected inputs.

3.5 Input delays and sampling time

The Hias process has significant time delays between the process inlet and outlet. The delay can be calculated as volume of ten basins divided by average wastewater flowrate $F_{average}$ as in Equation 6. The time delay is time variant, but in this work estimated as time invariant.

$$Td = \frac{V_{total}}{F_{average}} = \frac{10 \cdot 215m^3}{5.58m^3/min} = 385min \quad (6)$$

The six input variables consists of inlet water flowrate F, the inlet carbon composition COD, inlet nitrate and nitrite composition NOX and total suspended solids ΔTSS all four with a delay of 385min, aeration rate in basin 5 $FO5$ with delay of 270 min, and aeration rate in basin 8 $FO8$ with delay of 115 min. The data was imported to Matlab system identification toolbox with sampling time of 10 minutes.

3.6 Transfer functions

The estimated transfer function model parameters and error indices for estimation and validation data sets are given in Table 3 for TF0, a pure gain with a delay, and Table 4 for TF1, a first order model with delay. Time constants and delay are given in minutes. The signs of the gains K_p in the TF0 model are not quite as expected, but the modeling results are acceptable for estimation data (7%) and validation data (18%) and the model follows dynamic trends. The first order transfer function model follows the dynamic trends in the data as illustrated in Figure 3. The time constants Tp of the second order model TF1 4 are unacceptable large and give no physical interpretation, however the results for estimation and validation data are acceptable. We choose further to use the first order transfer function model TF0.

3.7 Time-series models

Time-series models in the format of auto-regressive with exogenous inputs were developed and tested. The input and output variables, and input delays were the same as for the transfer functions. The parameters and fitness index are shown in Table 5 for first order ARX110 model and in Table 6 for second order ARX220 model. The signs of the coefficients b_{i1} in the ARX110 model are not quite as expected, as coefficients for the aeration rates FO_i are positive. However the modeling results are very good for estimation data (69%) and acceptable for validation data (7%) and the model follows most of the dynamic trends in Figure 4. The results for the second order ARX model are similar to the first order model (70% and 7%), hence we work further with the first order model.

3.8 State-space models

State-space models were developed and tested. The input and output variables, and input delays, were the same as for the transfer functions. A first order (SS1) and a second order (SS2) state space model were identified. The parameters and results are shown in Table 7 for the SS1. The results for the second order state-space model were not successful. The signs of the coefficients b_{i1} in the SS1 model are logical for the phosphorus removal phenomena, except for aeration rates $FO5$ and $FO8$. Coefficient A for previous phosphorus measurements is ten fold compared to the input coefficients B, which implicates that the model is relies heavily on the previous effluent phosphorus measurements. As the data set is rather large, the parameter uncertainties for A, B and C are also very large. However, the modeling results are sufficient for estimation

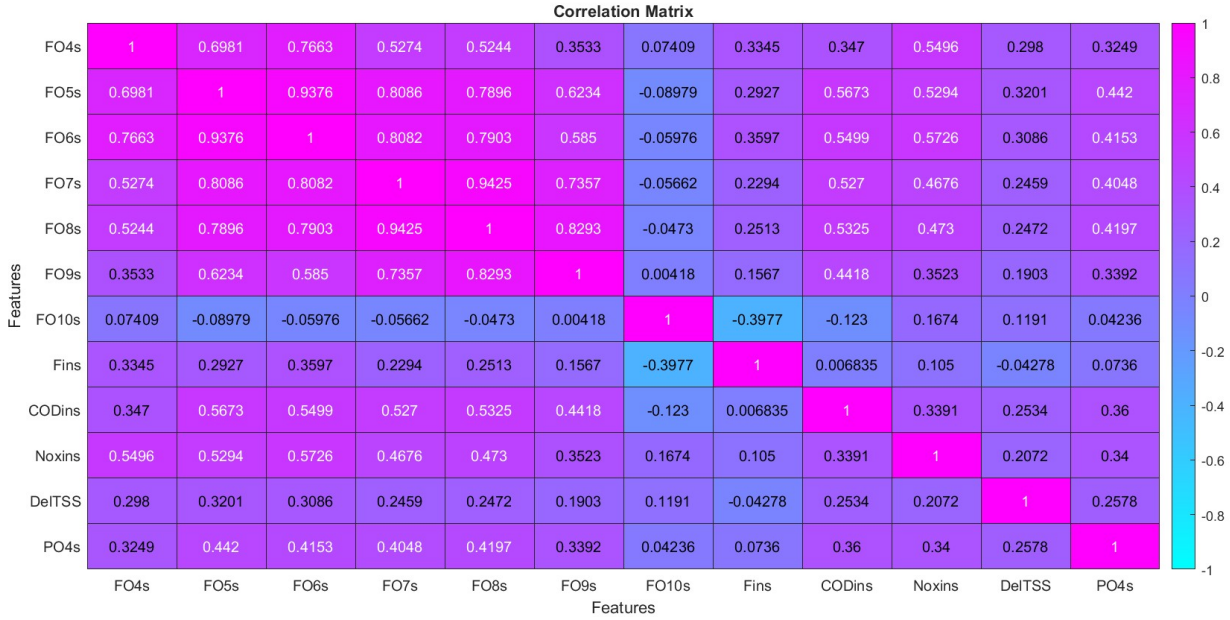


Figure 2. Heatmap between variables

data (6%) and good for validation data (18%) and the model follows dynamic trends in Figure 5. Hence, we work further with the first order state-space model.

Table 3. TF0 parameters

Input	Td	Kp
F	385	-6.68 ·10⁻2
COD	385	0.9857 ·10⁻2
NOX	385	8.989 ·10⁻2
FO5	270	2.317 ·10⁻2
FO8	115	7.627·10⁻2
ΔTSS	385	23.24 ·10⁻2
FITest	6.518	
FITval	17.67	

Table 4. TF1 parameters

Input	Td	Kp	Tp
F	385	0.293	327.5
COD	385	2.5	24838
NOX	385	0.24	159611
FO5	270	1.27	6.21e-7
FO8	115	2.6	20177
Δ TSS	385	-11.3	27322
FITest	15.89		
FITval	10.46		

Table 5. ARX110 parameters

Output	constant	a_1
$PO_{4,out}$	1	0.9497
Input	delay	b_{i1}
F	385	6.067 ·10⁻3
COD	385	4.392·10⁻3
NOX	385	0.6784·10⁻3
FO5	270	0.3634·10⁻3
FO8	115	4.21·10⁻3
ΔTSS	385	-3.26·10⁻3
FITest	68.85	
FITval	6.743	

Table 6. ARX220 parameters

Output	constant	a_1	a_2
$PO_{4,out}$	1	0.6529	0.3123
Input	delay	b_{i1}	b_{i2}
F	385	11.09 ·10⁻3	-3.206 ·10⁻3
COD	385	6.826 ·10⁻3	- 3.953 ·10⁻3
NOX	385	-2.044 ·10⁻3	+ 1.693 ·10⁻3
FO5	270	0.7614 ·10⁻3	- 0.7147 ·10⁻3
FO8	115	0.7298 ·10⁻3	+ 3.233 ·10⁻3
ΔTSS	385	-37.75 ·10⁻3	+ 43.62 ·10⁻3
FITest	70.42		
FITval	6.5		

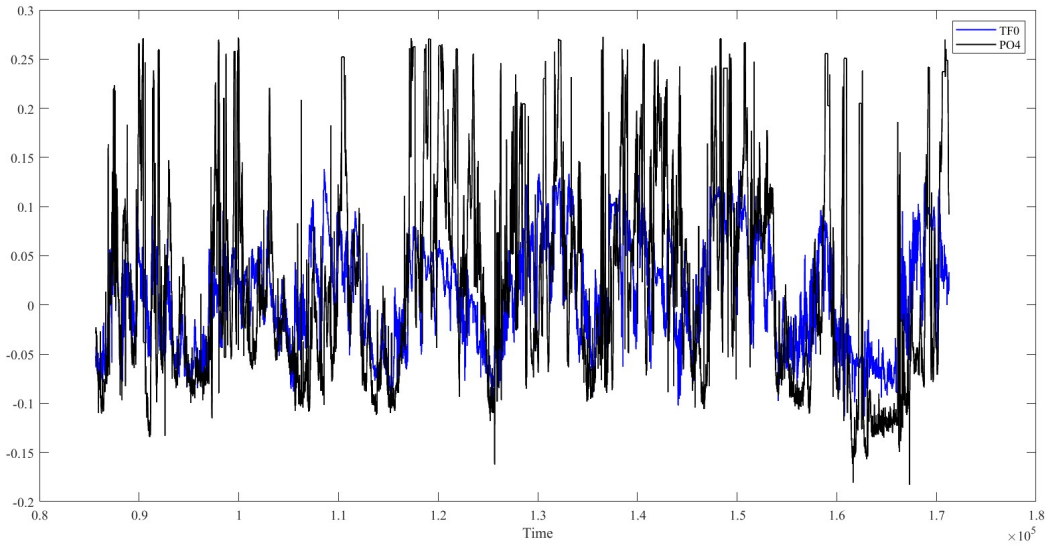


Figure 3. Scaled $PO_{4,out}$ measurement with black and TF0 estimate with blue, time in minutes.

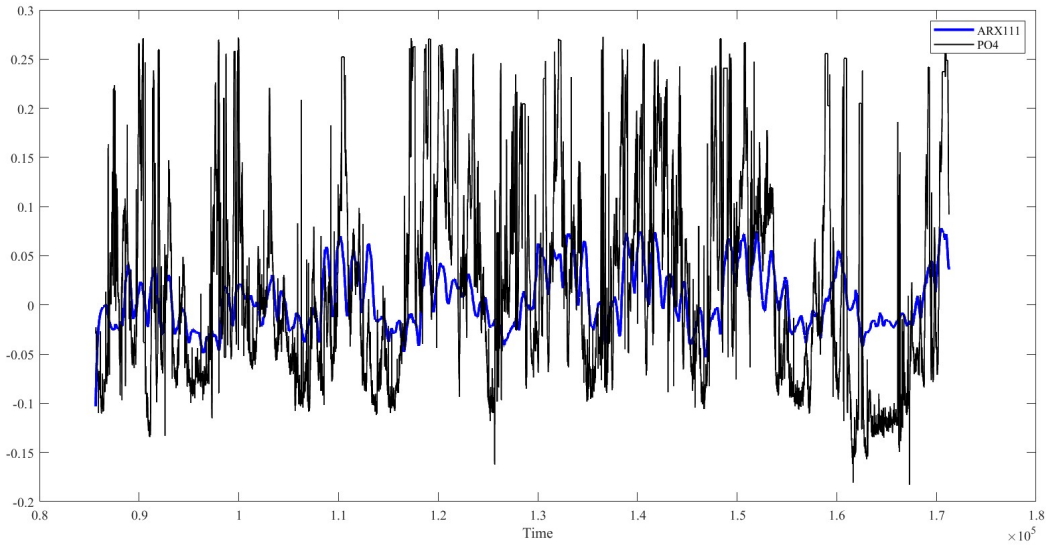


Figure 4. Scaled $PO_{4,out}$ measurement with black and ARX110 estimate with blue, time in minutes.

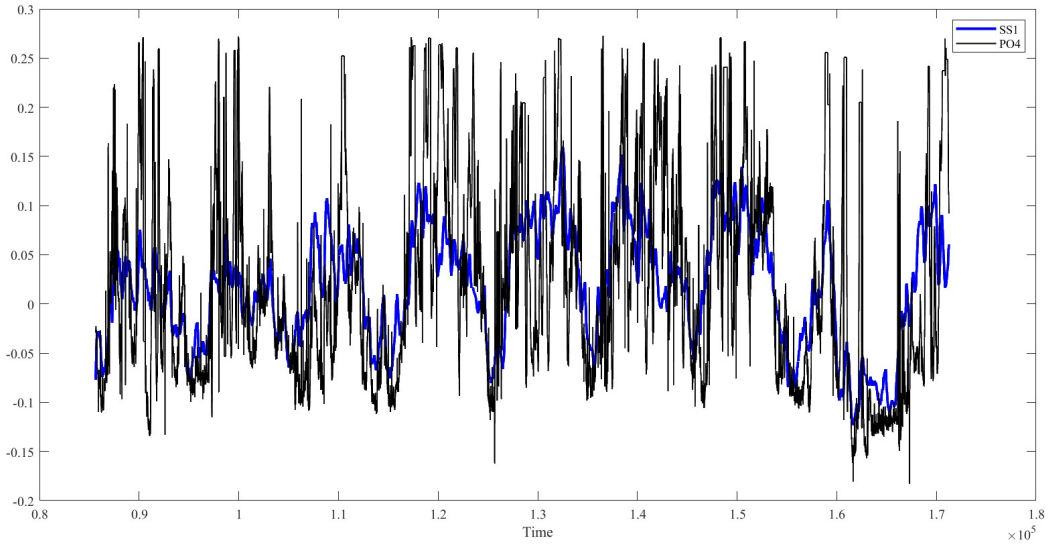


Figure 5. Scaled $PO_{4,out}$ measurement with black and SS1 estimate with blue, time in minutes.

Table 7. SS1 parameters

State	-	a_1
x(t)	-	-72.92 ·10⁻4
Inputs	delay	b_{i1}
F	385	-3.731 ·10⁻4
COD	385	0.256 ·10⁻4
NOX	385	1.548 ·10⁻4
FO5	270	0.139 ·10⁻4
FO8	115	0.973 ·10⁻4
ΔTSS	385	-1.624 ·10⁻4
Output	-	c_1
$PO_{4,out}$	-	7.578
FITest	6.178	
FITval	18.3	

4 Discussion and summary

Three data-driven modeling methods were tested for the four months data set from Hias municipal water resource recycling facility. Selection of input variables and determination of the input delays had most effect on the modeling accuracy. Using all the aeration rates and inlet flowrate, COD and NOX with the fitness index was around 1 % and the model was vaguely following the dynamics in the data (Nermo, 2023). Introducing the storage capacity estimated with difference of total suspended solids ΔTSS and reducing number of input variables (aeration rates FO_i) and fixing the input delay increased the fitness index significantly. We tried to reduce the number of inputs further by omitting the other aeration rate FO_8, but this did not improve the results.

The model fitness to estimation data was best with time-series models, around 70%. The model fitness to validation data was best with state-space models, 18%. All the first order models follow the dynamic changes in the data. The model parameter signs did not have quite the logical interpretation to the biological phosphorus removal phenomena between the input and output variables. As the data set is rather large, the parameter uncertainty is large, and an estimation method updating the model output, such as Kalman filter, could be developed for the online application. The answer to our research question is: state-space model can reproduce nutrient variations in the Hias Process with sufficient accuracy. Hence, we work further with the first order state space model and prepare an online application estimating the effluent nutrient concentration.

Further work

Further work will encompass work on data preprocessing methods suitable for online use. We will test different input variables sets, possibly excluding $FO8$ and including dissolved oxygen $DO5$ as state variable for state-space model. With this model we will develop and test novel control strategies for the Hias Process.

Further work is suggested on developing sub-models for the Hias Process that estimate phosphorus $PO_{4,out}$ concentration in basins with newly installed conductivity and redox-potential measurements. Potentially these virtual measurements of phosphorus in the process can then be used as inputs for the outlet phosphorus modeling.

5 Acknowledgements and contributions

RFF Innlandet, Norway, is gratefully acknowledged for funding the PACBAL project (nr 337727).
Fruitful discussions with Oulu University researchers Jani Tomperi, Henri Pörhö and Esko Juuso, and CEIWA research project, are gratefully acknowledged for inspiration on model types and input variable selection.
Torgeir Saltnes at Hias How2O, Katrine Marsten Jansen from Hias IKS water resource recycling facility and provided valuable insights in the Hias Process. Tobias Korten (Hias IKS) and Katrine Marsten Jansen have supported data collection. Axel Tveiten Bech and Ola Solli Grønningsæter from Digitread Connect have implemented IoT platform and performed online data collection. Anders Øfsti (Hias How2O) and Katrine Marsten Jansen (Hias IKS) have reviewed the article.
Professor Tiina Komulainen is the principal investigator of the project. She has supervised modeling work, provided the initial idea on input-output design and data-driven model types and structures. She has also prepared the article draft. Master student Einar Nermo done initial work on identifying transfer functions. Research assistant A.Malik Baqeri has continued the system identification work, improved the data preprocessing methods and identified the data-driven models under Professor Komulainen's supervision. Professor Olga Korostynska and Associate Professor Arvind Keprate have co-supervised Einar's master project. Professor Korostynska has done the critical review of this article.

References

El-Rawy, M., Abd-Ellah, M. K., Fathi, H., & Ahmed, A. K. A. (2021). Forecasting effluent and performance of wastewater treatment plant using different machine learning techniques. *Journal of Water Process Engineering*, *44*, 15. doi: https://-doi.org/10.1016/j.jwpe.2021.102380

EuropeanCommission. (2021). *Eu strategy on energy system integration.* Retrieved from `https://energy.ec.europa.eu/topics/energy-system-integration/eu-strategy-energy-system-integration_en`

EuropeanCommission. (2022). *Proposal for a revised urban wastewater treatment directive.* Retrieved from `https://environment.ec.europa.eu/publications/proposal-revised-urban-wastewater-treatment-directive_en`

Henze, M., Gujer, W., Mino, T., Matsuo, T., Wentzel, M. C., v.R. Marais, G., & Loosdrecht, M. C. V. (1999). Activated sludge model no.2d, asm2d. *Water Science Technology*, *39*(1), 165-182.

Ljung, L. (1999). *System identification - theory for the user* (2nd ed.). Prentice Hall. Retrieved from `https://onlinelibrary.wiley.com/doi/10.1002/047134608X.W1046`

Ly, Q. V., Truong, V. H., Ji, B., Nguyen, X. C., Cho, K. H., Ngo, H. H., & Zhang, Z. (2022). Exploring potential machine learning application based on big data for prediction of wastewater quality from different full-scale wastewater treatment plants. *Science of the Total Environment*, *832*, 12. doi: http://dx.doi.org/10.1016/j.scitotenv.2022.154930

Nair, A. M., Fanta, A., Haugen, F. A., & Ratnaweera, H. (2019). Implementing and extended kalman filter for estimating nutrient composition in a sequential batch mbbr pilot plant. *Water Science Technology*, *80*(2), 317–328. doi: https://-doi.org/10.2166/wst.2019.272

Nair, A. M., Gonzales-Silva, B. M., Haugen, F. A., Ratnaweera, H., & Østerhus, S. W. (2020). Real-time monitoring of enhanced biological phosphorus removal in a multistage ebpr-mbbr using a soft-sensor for phosphates. *Journal of Water Process Engineering*, *37*(101494), 13. doi: https://-doi.org/10.1016/j.jwpe.2020.101494

Nermo, E. (2023). *Mpc of nutrient removal in wastewater treatment process* [Master Thesis]. Oslo Metropolitan University.

Rudi, K., Goa, I. A., Saltnes, T., Sørensen, G., Angell, I. L., & Eikås, S. (2019). Microbial ecological processes in mbbr biofilms for biological phosphorus removal from wastewater. *Water Science Technology*, *79*(8), 1467–1473. doi: https://-doi.org/10.2166/wst.2019.149

Tomperi, J., & Leiviskä, K. (2018). Comparison of modelling accuracy with and without exploiting automated optical monitoring information in predicting the treated wastewater quality. *Environmental Technology*, *39*(11), 1442–1449. doi: https://-doi.org/10.1080/09593330.2017.1331267

Banks of Gaussian Process Sensor Models for Fault Detection in Wastewater Treatment Processes

H.L. Ivan[a] J.A. Ivan[b]

[a]*Mälardalen University, Sweden*
[b]*Örebro University, Sweden*
[a]heidi.ivan@mdu.se, [b]jean-paul.ivan@oru.se

Abstract

The harsh operating environment in a wastewater treatment process (WWTP) makes sensor faults commonplace. Detecting these faults can be challenging due to the complex process dynamics, unknown inputs, and general noise in the process and measurements. Comparing sensor readings against predictions from a physics-based or data-driven model of the WWTP is a common strategy for detecting such faults. In this work sensor measurements are directly modelled using Gaussian process (GP) regression, a data-driven multivariate approach. These GP sensor models are, with a generalised product of experts, combined into a dedicated fault isolation scheme resembling traditional observer bank methods. The residuals are monitored with a multivariate exponentially weighted moving average chart which is used for fault detection and isolation. The method is evaluated using simulated data generated with the Benchmark Simulation Model No. 1 WWTP. Fault detection performance is reported using several standard metrics such as false alarms, missed detections, time to detection, and successful fault isolations, with emphasis on reporting across a wide range of sensors and faults to provide a point of comparison for future studies. The proposed approach performs well across these metrics. Given sufficient data representative of normal operation, this approach can easily be adapted across a wide variety of plant configurations and can be used to create operator-friendly diagnostics resembling classical control charts.

1 Introduction

Wastewater treatment plants (WWTPs), like many critical components of public infrastructure, are gradually shifting to higher levels of automation in process operation. This is in part driven by global incentives to shift towards water resource recovery facilities, in conjunction with progress in regulation of the environmental impact of WWTPs, and typical cost incentives of reducing energy and material consumption. However, as in any process, automation depends on reliable process supervision; this is particularly challenging in WWTPs as sensors are often sparse and subject to harsh operating conditions.

A key task in a process supervision system is fault detection (FD). A common model-based FD strategy is to generate a residual signal from the difference between model predictions and actual sensor measurements (Chen & Patton, 1999). Predictions can stem from physics-based or data-driven models. Data-driven methods have risen in popularity as they generally require less extensive process-specific knowledge. However, forgoing process-specific knowledge means that more data is required for fitting data-driven models. Moreover, data-driven methods - neural networks as an archetypal example - can exhibit out-of-distribution overconfidence and in-distribution sensitivity to adversarial examples (Szegedy et al., 2014).

Gaussian processes (GPs) are a class of data-driven models which explicitly model uncertainty, and provide clear avenues - see for example (Jidling et al., 2017) - for reintroducing domain knowledge into learned models (Rasmussen et al., 2006). This principled treatment of uncertainty in the process model is useful as typically the residual generation process is complicated by requirements to reject model uncertainty and process disturbances while remaining sensitive to faults (Witczak, 2007).

Fault isolation (FI) - which requires FD - also requires further structure in the generated residuals. Two such structures are common: dedicated schemes - wherein each residual in a set is only *sensitive* to a single fault - and generalised schemes - wherein each residual is *insensitive* to only a single fault (Witczak, 2007; Chen & Patton, 1999). In WWTPs these schemes have previously been applied using physics-based state estimators configured in banks of observers (Nejjari et al., 2008; Nagy-Kiss & Schutz, 2013). However, the complexity of the process makes data-driven state estimation attractive - as in other fields (Palma et al., 2005; Sina Tayarani-Bathaie & Khorasani, 2015).

An issue which arises in using these schemes for FI is that the sets of residuals that need to be monitored for FD become large (one set per observer). The use of

multivariate statistical process monitoring techniques, such as the Hotelling T^2 chart (Hotelling, 1947) and the multivariate exponentially weighted moving average (MEWMA) chart (Lowry et al., 1992), can alleviate some of these difficulties. The latter is often more sensitive to smaller faults and slow drift faults (Montgomery, 2009). Creating a single, interpretable, and easily visualisable FD statistic for monitoring is of high priority in the WWTP industry; which is traditionally dominated by operator expertise.

In this work we illustrate the feasibility of the use of GP regression based sensor models combined using a generalised product of experts into an dedicated FI scheme for sensor fault detection in a standard biological WWTP. The ability to detect sensor faults on both controlled variable sensors, and ordinary monitored variables, is shown for two fault profiles and varying fault sizes and durations. The diagnostic performance is based on the number of faults detected, time to detection, number and duration of false alarms, as well as the number of faults correctly isolated.

2 Background

This section describes relevant ideas and theoretical perquisites to clarify the method section. In §2.1 FI schemes in general and the modification using direct sensor models proposed in this work are described. §2.2 concerns GP regression; used to create the aforementioned sensor models. Thereafter §2.3 covers products of experts, which combine several GP models into a dedicated FI scheme. Finally, §2.4 describes the MEWMA chart used for monitoring the residuals.

2.1 Dedicated FI Schemes without State Estimation

In a typical dedicated FI scheme residuals are generated from a bank of state estimators, where each estimator in the bank is ignorant of one of the sensors (Chen & Patton, 1999; Witczak, 2007). This is shown in (1): a sequence of measurements $\boldsymbol{y}_{-i,1:t}$ from time 1 to t (where $\boldsymbol{y}_{-i,t}$ denotes the vector $[y_1,\ldots,y_{i-1},y_{i+1},\ldots,y_n]_t$) is used to estimate the state $\hat{\boldsymbol{x}}_t^{-i}$ from which a sensor model estimates the sensor measurements $\hat{\boldsymbol{y}}_t^{-i}$. The notation $-i$ in a superscript indicates a state/sensor estimate is ignorant of measurements from sensor i, whereas a subscript indicates a vector missing state/sensor i.

$$\boldsymbol{y}_{-i,1:t} \xrightarrow{\text{state estimation}} \hat{\boldsymbol{x}}_t^{-i} \xrightarrow{\text{sensor model}} \hat{\boldsymbol{y}}_t^{-i} \qquad (1)$$

This estimate is used to compute a residual between the sensor estimates and measurements for FD/FI.

However, performing this state estimation in the WWTP is often difficult. In response to this difficulty the feasibility of directly estimating $\hat{\boldsymbol{y}}_t^{-i}$ from $\boldsymbol{y}_{-i,1:t}$ is considered. However, if the sequence $\boldsymbol{y}_{-i,1:t}$ is assumed to be Markovian, the problem can be further simplified by instead considering the estimation of $\hat{y}_{i,t}^{-i}$ directly from $\boldsymbol{y}_{-i,t}$. There is a problem with this: in the scheme shown in (1) if there is a fault in sensor i this appear as a) a residual in $\hat{y}_{i,t}^{-i}$, and b) in all $\hat{\boldsymbol{y}}_t^{-j}$ for $i \neq j$. This asymmetry is what allows FI from the residuals in (1). Directly estimating $\hat{y}_{i,t}^{-i}$ from $\boldsymbol{y}_{-i,t}$ eliminates this property - a fault on sensor i affects every estimate. This can be remedied by repeating the leave-out-one pattern in the original scheme. In this work we propose using a bank of $n(n-1)$ models $\mathcal{M}^{-ij}$ where $i \neq j$ where each model estimates $\hat{y}_{i,t}$ from all sensors except *i and j*. Given the Markov assumption the time subscripts are omitted:

$$\boldsymbol{y}_{-ij} \xrightarrow[\mathcal{M}^{-ij}]{\text{sensor model}} \hat{y}_i^{-ij}. \qquad (2)$$

For each i this creates $n-1$ estimates $\hat{y}_i^{-ij}$, each insensitive to a fault in a different sensor $j \neq i$. This reestablishes the required asymmetry for FI.

2.2 Gaussian Process Regression

Gaussian process (GP) regression (Williams & Rasmussen, 1995; Rasmussen et al., 2006) is a non-parametric regression method which assumes the target function $f : \mathcal{Y}^{-ij} \to \mathbb{R}$ to be a stochastic (Gaussian) process and conditions this prior process on observations to obtain a posterior distribution over functions that explain the observations. A Gaussian distribution is fully specified by its mean vector and covariance matrix. Analogously, a GP is fully specified by a mean function $m^{-ij} : \mathcal{Y}^{-ij} \to \mathbb{R}$ and a covariance function $k_f^{-ij} : \mathcal{Y}^{-ij} \times \mathcal{Y}^{-ij} \to \mathbb{R}$. This is typically denoted $f \sim \mathcal{GP}(m^{-ij}, k_f^{-ij})$. Assuming the observations $y_i^{-ij}(\boldsymbol{y}_{-ij}) = f(\boldsymbol{y}_{-ij}) + \varepsilon$ of f are perturbed by Gaussian noise $\varepsilon \sim \mathcal{N}(0, \sigma_n^2)$, the measurement process is also Gaussian and is denoted,

$$y_i^{-ij} \sim \mathcal{GP}\left(m^{-ij}, k^{-ij}\right). \qquad (3)$$

Here, the measurement covariance k^{-ij} is a sum of the 'base' covariance of f and the noise of the observation process: $k^{-ij}(y_{-ij}, y'_{-ij}) = k_f^{-ij}(y_{-ij}, y'_{-ij}) + \delta_{yy'}\sigma_n^2$ where $\delta_{yy'}$ is the Kronecker delta on $y_{-ij} = y'_{-ij}$.

Like many forms of Bayesian inference, GP regression has historically been associated with heavy computational costs. However, frameworks such as GPyTorch (Gardner et al., 2018), taking advantage of modern hardware and theoretical progress, allow practical use of GPs with standard covariance functions.

2.3 Generalised Products of Experts

Taken in combination, §2.1 and §2.2 suggest using GP regression to learn a bank of models $\mathcal{M}^{-ij}$. Structure in this bank can be exploited for FI, but in order to

allow for clear visualisation and interpretation by operators the signals from these $n \times (n-1)$ models must ideally be used to generate a single FD statistic.
One method of obtaining a combined predictive distribution $p_c(y|x)$ from several GP posteriors $p^i(y|x)$ is the (generalised) product of experts (GPoE/PoE) (Cao & Fleet, 2015),

$$p_c(y|x) \;=\; \frac{1}{Z}\prod_i p^{\alpha_i(x)}(y|x), \tag{4}$$

where Z is a normalisation constant. The annealing parameters α_i are used to amplify or diminish the importance of each model's contribution to the combined distribution. The simplest parameters, corresponding to a PoE, $\alpha_i(x) = 1$, are used in this study. If each model in the product is Gaussian the combined distribution is also Gaussian with mean and covariance

$$m_c^{-i}(x) \;=\; k_c^{-i}(x,x)\sum_j m^{-ij}(x)\alpha_j(x)\lambda^{-ij}(x), \tag{5}$$

$$k_c^{-i}(x,x) \;=\; \left(\sum_j \alpha_j(x)\lambda^{-ij}(x)\right)^{-1}. \tag{6}$$

Where $\lambda^{-ij}(x) := 1/k^{-ij}(x,x)$. Combining the predictions from the bank in this way produces a combined posterior for each sensor,

$$y_i^{-i} \sim \mathcal{GP}(m_c^{-i}, k_c^{-i}). \tag{7}$$

The vector $\hat{\boldsymbol{y}} = [y_1^{-1}, \ldots, y_n^{-n}]^T$ denotes the full estimate of the sensor state obtained from the GPoE.

2.4 Multivariate Process Monitoring

The multivariate exponentially weighted moving average (MEWMA) chart, first proposed by Lowry et al. (1992), utilises information from successive samples and is therefore relatively sensitive to small shifts in the mean of the monitored variable. In this application, that is the standardised residual vector $\boldsymbol{r}_t$, where the raw residuals are $\tilde{\boldsymbol{r}}_t = \boldsymbol{y}_t - \mathbb{E}[\hat{\boldsymbol{y}}_t]$. The relevant parameters are defined as (Montgomery, 2009)

$$\boldsymbol{Z}_t \;=\; \lambda \boldsymbol{r}_t + (1-\lambda)\boldsymbol{Z}_{t-1} \tag{8}$$

where $0 \leq \lambda \leq 1$ and $\boldsymbol{Z}_0 = \boldsymbol{0}$. The statistic monitored on the chart is

$$T_t^2 \;=\; \boldsymbol{Z}_t^T \boldsymbol{\Sigma}_{\boldsymbol{Z}_t}^{-1} \boldsymbol{Z}_t \tag{9}$$

where

$$\boldsymbol{\Sigma}_{\boldsymbol{Z}_t} \;=\; \frac{\lambda}{2-\lambda}\left[1-(1-\lambda)^{2t}\right]\boldsymbol{\Sigma}. \tag{10}$$

$\boldsymbol{\Sigma}$ represents the covariance of $\boldsymbol{r}$ from a collection of samples when the process is known to be operating normally. The performance of the MEWMA chart is tuned by adjusting λ, the smoothing factor, as well as the limit H.
When $T_t^2 > H$ the limit is violated, indicating a fault. The source of the deviation can be determined by decomposition of the MEWMA statistic, described by VandenHul (2002). This requires recalculating T_t^2 for the value of t at which the limit is violated based on $\boldsymbol{r}^{-i} := [r_1, \ldots, r_{i-1}, r_{i+1}, \ldots, r_n]^T$, thus generating n values for $T_t^{2,-i}$. By observing which $T_t^{2,-i}$ decreases the most compared to T_t^2 the responsible residual can be isolated.

3 Methodology

The high-level strategy proposed for performing and evaluating FD/FI using banks of GPs is as follows. Data, with and without faults, is generated in simulation (§3.1) and used to train a bank of GP sensor models (§3.2). These models are combined in a GPoE, and the combined predictions used to generate a MEWMA chart on the residuals. FD/FI statistics are calculated over a set of 320 faults per sensor (12 total) in a typical sensor set - parameterised by fault type, size, duration, and start time.

3.1 Simulation

The Benchmark Simulation Model No.1 (BSM1) was used to simulate the operation of the WWTP. The simulation platform consists of two anoxic reactors of 2000 m^3 and three aerated reactors of 3999 m^3 followed by a secondary settler of 6000 m^3 (Gernaey et al., 2014). The BSM1 process contains two standard control loops: S_{NO} control in the second reactor with set-point of 1 $\mathrm{g\,N\,m^{-3}}$ and S_O control in the fifth reactor with set-point 2 $\mathrm{g\,O_2\,m^{-3}}$ (Gernaey et al., 2014).
Sensor measurements of dissolved oxygen (S_O), alkalinity (S_{ALK}), total suspended solids (TSS), nitrate/nitrite nitrogen (S_{NO}), and ammonium/ammonia nitrogen (S_{NH}), at several points in the process were recorded at 15-minute intervals. These are hereafter denoted: $S_{NH,1}$, $S_{NO,2}$, $S_{O,3}$, $S_{NO,3}$, $S_{NH,3}$, $S_{O,4}$, $S_{NH,4}$, TSS_4, $S_{O,5}$, $S_{NO,e}$, $S_{ALK,e}$, TSS_e where the subscripts denote that the measurements are taken in the indicated tank number (1 to 5) or in the effluent (e). The the two controller outputs, $u_{NO,2}$ and $u_{O,5}$ were also recorded. The sensors were selected based on the approach in Marais, Zaccaria, Ivan, & Nordlander (2022); Ivan (2023).
The simulations used the BSM1 dry weather influent file, simulating two weeks of operation. The first week of simulated data was used as training data for GP training (§3.2) and chart calibration (§2.4, §3.4). The second week was held-out for testing FD/FI, where data from the eighth day was used for chart tuning and faults were introduced starting in the ninth day. Two fault types, bias and drift, were used with

Table 1. Fault parameters used for testing performance. All 320 parameter combinations were tested.

Size	Bias [σ]	1.5	2	3	5
	Drift [μ/Day]	0.1	0.25	0.5	1
Direction		+	−		
Start Time [Day]		8.75	9.5	10.25	11
Duration [Day]		0.5	1	1.5	2	2.5

varying parameters shown in Table 1. All combinations of size-direction-start-duration were simulated for each fault type for a total of 320 faults per sensor. The fault sizes are specified in proportion to the standard deviation (bias faults) and mean (drift fault) of the sensor signal as determined from the training data. All measurements are also standardised w.r.t. these statistics, $y_{i,t} = (\tilde{y}_{i,t} - \text{avg}_t\, \tilde{y}_{i,t})/\,\text{std}_t\, \tilde{y}_{i,t}$ where $\tilde{y}$ denotes the raw measurements.

3.2 Individual GP Sensor Models

The simulated sensor measurements $\mathcal{D} = (\boldsymbol{y}_t)_{t=1}^T$ are split into training and test data as described in §3.1. In order to create a supervised learning problem with FI asymmetry (see §2.1), $n \times (n-1)$ training sets are derived from this data. Each model $\mathcal{M}^{-ij}$ thus has associated datasets

$$\mathcal{D}^{-ij} \;=\; \{(\boldsymbol{x}_j, y_i) : \boldsymbol{x}_j = \boldsymbol{y}_{-ij}, \boldsymbol{y} \in \mathcal{D}\}. \quad (11)$$

As the measurements are standardised (§3.1), a GP prior with zero mean is placed over all models,

$$y_i^{-ij} \;\sim\; \mathcal{GP}(0, k^{-ij}). \quad (12)$$

The covariance k^{-ij} for each model is a sum of a linear kernel with noise parameter ψ^{-ij} and a squared exponential kernel with independent lengthscales ϕ^{-ij} for each input dimension and scale parameter σ^{-ij}:

$$\begin{aligned} k^{-ij}(\boldsymbol{x}_j, \boldsymbol{x}'_j) \;&=\; \sigma^{-ij} k_{\text{SE}}^{-ij}(\boldsymbol{x}_j, \boldsymbol{x}'_j) \\ &\quad + k_{\text{LIN}}^{-ij}(\boldsymbol{x}_j, \boldsymbol{x}'_j) + \sigma_n^2 \delta_{jj'}, \quad (13) \\ k_{\text{SE}}^{-ij}(\boldsymbol{x}_j, \boldsymbol{x}'_j) \;&=\; \exp\left(-\frac{1}{2}\boldsymbol{d}_j^T \Phi_{-ij}^{-2} \boldsymbol{d}_j\right), \quad (14) \\ k_{\text{LIN}}^{-ij}(\boldsymbol{x}_j, \boldsymbol{x}'_j) \;&=\; \psi^{-ij} \boldsymbol{x}_j^T \boldsymbol{x}'_j, \quad (15) \end{aligned}$$

where $\boldsymbol{d}_j = \boldsymbol{x}_j - \boldsymbol{x}'_j$ and $\Phi_{-ij} := \text{diag}\, \phi^{-ij}$. The initial covariance parameters are shown in Table 2. These GPs were defined using GPyTorch (Gardner et al., 2018) and each set of hyperparameters was independently optimised until convergence with respect to the marginal log-likelihood of each model. Optimisation was performed using the ADAM algorithm (Kingma & Ba, 2017) with learning rate 0.1 and learning rate decay 0.99.

Table 2. Initial covariance parameters and optimisation constraints.

Param.	Initial Value(s)	Optimisation Bounds
σ^{-ij}	1.0	$(0, 10)$
φ^{-ij}	Random $\in (0.5, 1.5)$	$(0, 10)$
ψ^{-ij}	1.0	$(0, 10)$
σ_n^2	1.0	$(0.01, \infty)$

3.3 Sensor Residuals via PoE

The GP posteriors are composed in a PoE, obtaining a combined estimate y_i^{-i} for each sensor:

$$\begin{aligned} y_i^{-i}(\boldsymbol{y}_{-i}) \;&\sim\; \mathcal{N}(m_c^{-i}, k_c^{-i}), \quad (16) \\ m_c^{-i}(\boldsymbol{y}_{-i}) \;&=\; k_c^{-i}(\boldsymbol{y}_{-i}, \boldsymbol{y}_{-i}) \sum_j m_\lambda^{-ij}(\boldsymbol{y}_{-ij}), \quad (17) \\ k_c^{-i}(\boldsymbol{y}_{-i}, \boldsymbol{y}_{-i}) \;&=\; \left(\sum_j \lambda^{-ij}(\boldsymbol{y}_{-ij})\right)^{-1}, \quad (18) \end{aligned}$$

where $m_\lambda^{-ij}(\boldsymbol{y}_{-ij}) := m^{-ij}(\boldsymbol{y}_{-ij}) \lambda^{-ij}(\boldsymbol{y}_{-ij})$. The model residuals are obtained from the PoE output at each timestep, $\hat{\boldsymbol{y}}_t$, as described in §2.4.

3.4 MEWMA Chart

As in §2.4, the residuals are standardised based on the training data:

$$r_{t,i} \;=\; (\tilde{r}_{t,i} - \text{avg}_t\, \tilde{r}_{t,i})/\,\text{std}_t\, \tilde{r}_{t,i}. \quad (19)$$

The covariance of the standardised residuals $\boldsymbol{\Sigma}$ was used to calibrate the chart according to (10). Small values for the smoothing factor, $\lambda = 0.05, 0.1, 0.2$, were evaluated; which in principle allow for the detection of smaller faults (Montgomery, 2009). For each value of λ the mean and standard deviation of the T^2 statistic was calculated during day 8 of the simulation data (§3.1) to determine an appropriate limit size H. Three values were tested for the limit size $H = \text{avg}_t\, T^2 + h\,\text{std}_t\, T^2$, for $h = 2, 3, 4$.
Charts with each combination of (λ, h) were used to monitor the performance of the process using (8) and (9). Every limit violation was treated as a fault alarm. In the case of correct fault detections, isolation was performed by constructing reduced charts of the persenor T^2-statistic, $T^{2,-i}$, as described in §2.4.

3.5 Diagnostics

As described in §2.4, a fault is detected when the chart limit H is crossed by the MEWMA T^2 statistic during the fault. For each pair of chart limit and smoothing factor the following statistics were recorded for all faults in Table 1: a) correct violation of H during a fault - fault detection (FD), b) number of incorrect

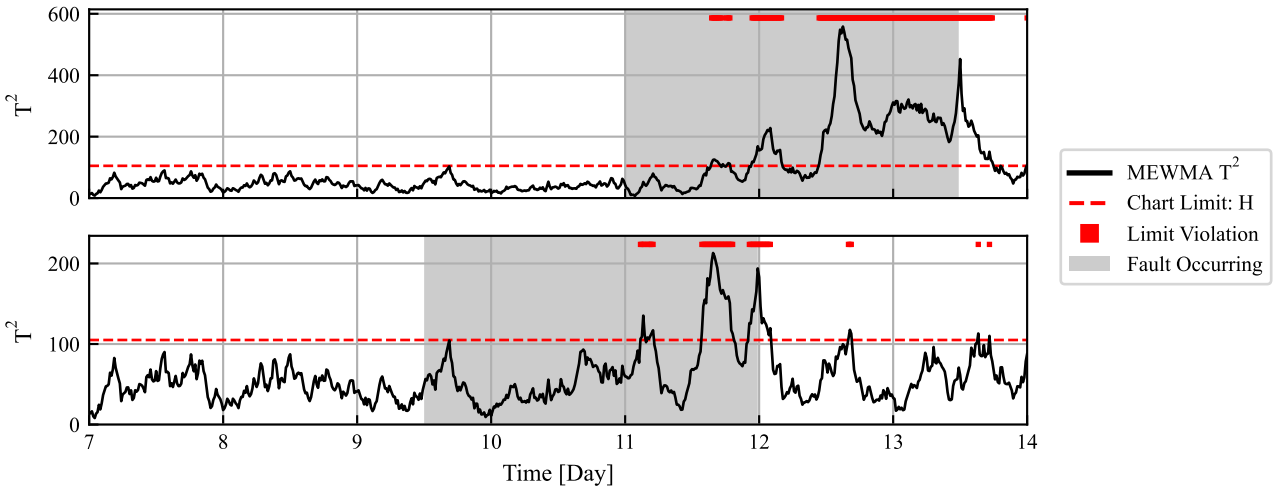


Figure 1. Illustration of the MEWMA chart showing the chart limit, the statistic, and limit violations. The time period during which the fault is occurring is highlighted. Top: Drift fault on $S_{O,5}$ with a rate of -0.25, Bottom: Drift fault on $S_{O,5}$ with a rate of -0.1.

chart limit violations - false alarms (#FAs), c) duration of false alarms (FA), d) time taken to detect the fault (TTD), and e) successful fault isolation (FI).

The MEWMA chart requires some time to return to normal after a fault stops. As such time spent above the chart limit immediately following a successful fault detection is not reported in the FA statistic. Note, defining a detection by *crossing* of the limit means that a false alarm preceding a fault which continues into the start of the fault does *not* constitute a detection.

Fault isolation was performed based on the mean of the T^2-decomposition during the first hour after violation. Only the $T^{2,-i}$ deviating most from T^2 was used for isolation. Isolation of faults of the controlled variable sensors was performed by monitoring the controller outputs, not the sensor measurements. For a discussion of 'fault hiding' on controlled variables see Marais, Zaccaria, & Odlare (2022).

4 Results and Discussion

Two MEWMA charts are shown in Figure 1 for two different drift faults on the $S_{O,5}$ sensor. It is clear that the smaller fault is harder to detect, shown by the longer detection time and the smaller values of T^2 relative to those of normal operation. Natural variation in the residuals, and therefore the T^2, can worsen the situation. For example, the T^2 statistic is low around day 10 - faults occurring near this point will be harder to detect due to the statistic being below its mean. This may be improved by reducing the nominal variance of the chart, requiring improved sensor estimates.

Overall the chart is clear and provides a good starting point for operator-friendly FD. With regard to FI, Figure 2 shows an example of an isolation plot, which could be shown to operators continuously using a rolling window on the decomposed T^2 statistic. The isolation chart shows clearly which residuals are contributing to variations in the MEWMA chart.

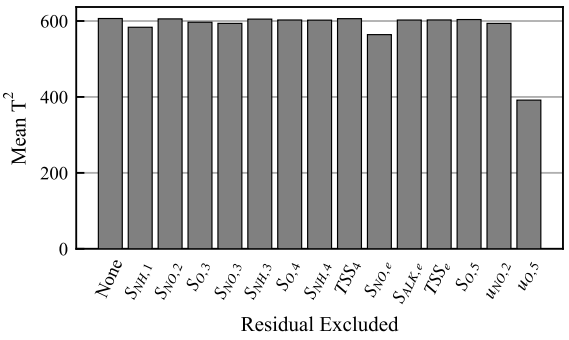


Figure 2. Example of an isolation plot for the drift fault in Figure 1-Top showing the mean of $T^{2,-i}$ (residual i excluded from T^2) during the isolation period. "None" denotes T^2: no residual excluded.

Figure 2 shows the mean value of $T^{2,-u_{O,5}}$ (i.e. the T^2 value calculated excluding $u_{O,5}$, the controller output for the $S_{O,5}$ controller) during the isolation period has the lowest value. This indicates the violation can be attributed to $u_{O,5}$, and therefore $S_{O,5}$.

The following sections present a more detailed analysis across chart parameters; fault types, sizes, and durations; and across different sensors.

4.1 Impact of MEWMA Chart Parameters

Broadly, the different values of (λ, h) affect FD/FI in accordance with theoretical expectations. As the limit size, h, increased detection becomes slower and less consistent, false alarms decrease, but FI becomes easier. As the smoothing factor, λ, increases the opposite occurs; smaller faults become detectable, but FI on these faults is more difficult, and false alarms increase. Small drift faults, in particular, are most sensitive to the change in the smoothing factor. These results are summarised in Table 3.

Table 3 clearly shows the expected trade-off that must

Table 3. Summary of diagnostics performance parameters, averaged over all fault types and characteristics, for the different chart parameters.

λ	h	FD [%]	FI [%]	TTD [d]	FA [d]	#FAs
0.20	2	0.98	0.69	0.16	0.33	15.31
0.20	3	0.90	0.82	0.28	0.07	4.64
0.20	4	0.87	0.85	0.30	0.01	1.13
0.10	2	0.95	0.70	0.22	0.24	11.07
0.10	3	0.90	0.80	0.28	0.04	2.88
0.10	4	0.87	0.85	0.31	0.00	0.14
0.05	2	0.94	0.72	0.24	0.30	7.04
0.05	3	0.92	0.79	0.27	0.08	5.04
0.05	4	0.89	0.83	0.31	0.02	1.16

be made in the MEWMA chart design: improved detectability comes at the expense of isolability and false alarms. For a given smoothing factor, detection rates decrease by between 5 % to 11 % and isolation rates increase by 15 % to 23 % as the limit size is increased. Detection times increase by 30 % to 46 % while the number and duration of false alarms decreases by 83 % to 97 %. Smaller smoothing factors are less sensitive to the limit size.

It is worth noting that a real FD system can reasonably run several combinations of chart parameters with the strengths and weaknesses of each chart in mind. Balancing these trade-offs, the remaining analysis proceeds with $(\lambda, h) = (0.1, 3)$.

4.2 Performance of Diagnostics

Table 4 shows a comparison between bias and drift type faults, averaged across all fault parameters and sensors. The false alarms are not included as they do not differ from those presented in Table 3; false alarms are chart-dependent, not fault-dependent.

The drift faults are, as expected, harder to detect and require a longer time on average before the faults are detected. However, the isolation of drift faults is not substantially lower than that of bias faults.

The relative difficulty of detecting and isolating faults in different sensors can be seen in Figure 3 where, averaged over all fault parameters, detection and isolation statistics are shown. The most challenging faults to isolate occur in the controlled variable sensors, that is $S_{NO,2}$ and $S_{O,5}$, the lowest average detection rate among all sensors also occurs in the former. This is expected: the controller works to keep these sensor values at the set-point, obfuscating the effects of the sensor faults on the sensor itself. The proposed residual scheme relies on the use of the controller output, as mentioned previously, to reliably circumvent this issue.

Figure 3 also shows that faults in sensors TSS_4, $S_{ALK,e}$, and $S_{NO,e}$ have some of the highest detection and isolation rates and shortest detection times. This is of special importance as sensors in the effluent are important for monitoring limits related to environmental regulations. In general, sensors which have high isolation rates, such as $S_{ALK,e}$ and $S_{NO,e}$, should be subject to further careful monitoring as it is possible that they are often the target of an incorrect isolation. In the faults tested these two sensors were responsible for 26 % of incorrect isolation cases.

In order to evaluate the effects of different fault sizes and durations on detectability and isolability, the results for a single sensor ($S_{NO,2}$ - a controlled variable) are shown in Figure 4.

Considering the bias faults first: all the faults are detected, and as the size of the fault increases the time to detection decreases to a minimum of 0.026 d, or 37 min. It might be expected that the isolation rate increase with the size of the fault, however, it is important to note that this fault occurs on a controlled variable sensor. This type of fault impacts the operation of the entire process through the control system, therefore, larger faults can have a larger impact on other process variables. This can make these faults more challenging to isolate as they disrupt other variables in the process.

The chaotic behaviour of the smallest drift fault likely has similar explanation - slow drift is corrected by the controller and propagates non-linearly throughout the system. Apart from this exception, the behaviour of the drift faults is unsurprising: larger faults are detected more reliably and more rapidly, and when a fault persists for longer it is both easier to detect and easier to isolate. The minimum detection time of the drift faults is around 0.44 d, or 10 h.

Comparing the results for faults on the $S_{NO,2}$ sensor to those in Marais, Zaccaria, & Odlare (2022), where a univariate EWMA chart is used, the detection times for the bias faults are slightly longer but the time for drift fault detection has been decreased by several hours. The number of false alarms are in the same order of magnitude, and the detection rate of drift faults has increased from 56 % to 64 % to between 75 % and 100 % for faults longer than 1 d. In Marais, Zaccaria, & Odlare (2022) the results were not broken down by duration of fault so this comparison is not exhaustive. Further comparisons with the broader literature are difficult due to inconsistencies in how results are reported, varying fault sizes, and incomparable plant configurations. A cursory comparison with Luca et al. (2021, 2023) shows detection times in similar ranges with possibly better performance on the bias faults.

5 Conclusions and Recommendations

Direct modelling of sensors using GP regression in a dedicated residual scheme and monitoring with a MEWMA chart can be used for FD/FI in a WWTP. Clear comparison with the broader literature is diffi-

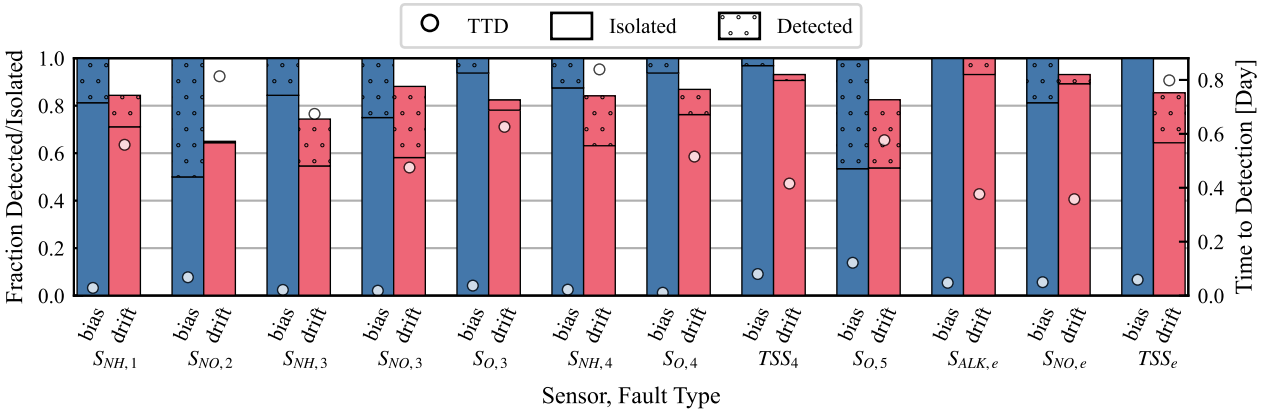


Figure 3. Detection and isolation rates, and time to detection split across bias and drift faults for each individual sensor.

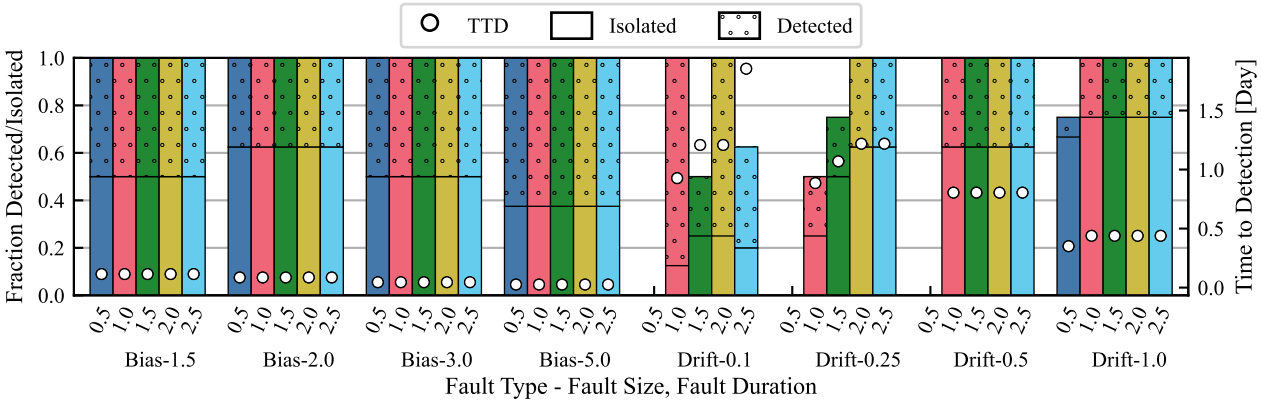


Figure 4. Detection and isolation rates, and time to detection across all faults on the $S_{NO,2}$ sensor. The bars represent a combination of fault type, size, and duration, and are grouped according to type and size.

Table 4. Detection and isolation statistics by fault type.

	FD [%]	FI [%]	TTD [d]
Bias	100	83	0.047
Drift	80	77	0.564

cult, as standardised reporting of performance evaluation parameters in studies performed in this field is lacking. In response to this difficulty, testing on a wide range of faults across many standard sensors has been reported in the hopes of facilitating future comparisons.

The method improves over a previous study using a univariate approach, and the results are comparable to other multivariate methods for FD/FI. Critically, the proposed approach is easy to visualise; a priority when developing FD/FI methods for an industry that relies heavily on operator expertise and shies away from uninterpretable automation.

The proposed approach leaves a great deal of room for further study. Without methodological changes, results across the each tested sensor can be documented, performance on out-of-distribution test data such as the BSM1 wet weather influent data can be evaluated, and more detailed FI studies carried out. The sensor models themselves can likely be simplified and made more interpretable by sharing parameters across models. Annealing the GPoE distributions, directly using the pre-GPoE sensor models in a generalised scheme, or other similar modifications to the sensor models could also yield improvements.

Acknowledgement

The authors acknowledge and express their gratitude to Dr Ulf Jeppsson and other IWA Task Group members for the availability of the BSM1 code. The work of J.A. Ivan was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

CRediT Author Contribution Statement

Conceptualization, Validation, Formal analysis, Investigation, Resources, Data curation, Writing: H.L.I., J.A.I.; Methodology: H.L.I. - FD/FI charts, J.A.I. - GP/GPoE models; Software: H.L.I. - BSM1/FD/FI, J.A.I. - GP/GPoE models; Visualization: H.L.I.;

References

Cao, Y., & Fleet, D. J. (2015). Generalized Prod-

uct of Experts for Automatic and Principled Fusion of Gaussian Process Predictions. *arXiv preprint arXiv:1410.7827*.

Chen, J., & Patton, R. J. (1999). *Robust Model-Based Fault Diagnosis for Dynamic Systems*. Springer US. doi: 10.1007/978-1-4615-5149-2

Gardner, J. R., Pleiss, G., Bindel, D., Weinberger, K. Q., & Wilson, A. G. (2018). Gpytorch: Blackbox matrix-matrix gaussian process inference with gpu acceleration. In *Advances in neural information processing systems*.

Gernaey, K. V., Jeppsson, U., Vanrolleghem, P. A., & Copp, J. B. (Eds.). (2014). *Benchmarking of Control Strategies for Wastewater Treatment Plants*. IWA Publishing. doi: 10.2166/9781780401171

Hotelling, H. (1947). Multivariate Quality Control Illustrated by Air Testing of Sample Bombsights. In C. Eisenhart, M. Hastay, & W. Wallis (Eds.), *Techniques of Statistical Analysis* (pp. 111–184). New York: McGraw Hill.

Ivan, H. L. (2023). *Fault Detection in Wastewater Treatment : Process Supervision to Improve Wastewater Reuse* (Licentiate dissertation). Mälardalen University, Västerås, Sweden.

Jidling, C., Wahlström, N., Wills, A., & Schön, T. B. (2017). Linearly constrained Gaussian processes. In *Advances in Neural Information Processing Systems* (Vol. 30). Curran Associates, Inc.

Kingma, D. P., & Ba, J. (2017). Adam: A Method for Stochastic Optimization. *arXiv preprint arXiv:1412.6980*.

Lowry, C. A., Woodall, W. H., Champ, C. W., & Rigdon, S. E. (1992). A Multivariate Exponentially Weighted Moving Average Control Chart. *Technometrics*, *34*(1), 46–53. doi: 10.2307/1269551

Luca, A.-V., Simon-Várhelyi, M., Mihály, N.-B., & Cristea, V.-M. (2021). Data Driven Detection of Different Dissolved Oxygen Sensor Faults for Improving Operation of the WWTP Control System. *Processes*, *9*(9), 1633. doi: 10.3390/pr9091633

Luca, A.-V., Simon-Várhelyi, M., Mihály, N.-B., & Cristea, V.-M. (2023). Fault Type Diagnosis of the WWTP Dissolved Oxygen Sensor Based on Fisher Discriminant Analysis and Assessment of Associated Environmental and Economic Impact. *Applied Sciences*, *13*(4), 2554. doi: 10.3390/app13042554

Marais, H. L., Zaccaria, V., Ivan, J.-P. A., & Nordlander, E. (2022). Detectability of Fault Signatures in a Wastewater Treatment Process. In *The First SIMS EUROSIM Conference on Modelling and Simulation, SIMS EUROSIM 2021, and 62nd International Conference of Scandinavian Simulation Society, SIMS 2021* (pp. 418–423). Virtual Conference, Finland. doi: 10.3384/ecp21185418

Marais, H. L., Zaccaria, V., & Odlare, M. (2022). Comparing statistical process control charts for fault detection in wastewater treatment. *Water Science and Technology*, *85*, 1250–1262. doi: 10.2166/wst.2022.037

Montgomery, D. C. (2009). *Introduction to statistical quality control* (6th ed ed.). John Wiley & Sons, Incorporated.

Nagy-Kiss, A. M., & Schutz, G. (2013). Estimation and diagnosis using multi-models with application to a wastewater treatment plant. *Journal of Process Control*, *23*(10), 1528–1544. doi: 10.1016/j.jprocont.2013.09.027

Nejjari, F., Puig, V., Giancristofaro, L., & Koehler, S. (2008). Extended Luenberger Observer-Based Fault Detection for an Activated Sludge Process. *IFAC Proceedings Volumes*, *41*(2), 9725–9730. doi: 10.3182/20080706-5-KR-1001.01645

Palma, L., Coito, F., & da Silva, R. (2005). Process fault diagnosis approach based on neural observers. In *2005 IEEE Conference on Emerging Technologies and Factory Automation* (Vol. 1, pp. 4 pp.–1060). doi: 10.1109/ETFA.2005.1612642

Rasmussen, C. E., Williams, C. K. I., & Bach, F. (2006). *Gaussian Processes for Machine Learning*. MIT Press.

Sina Tayarani-Bathaie, S., & Khorasani, K. (2015). Fault detection and isolation of gas turbine engines using a bank of neural networks. *Journal of Process Control*, *36*, 22–41. doi: 10.1016/j.jprocont.2015.08.007

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., & Fergus, R. (2014). Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199*.

VandenHul, S. P. (2002). *Decomposition of the MEWMA statistic* (Ph.D.). University of Northern Colorado, United States. (ISBN: 9780493758718)

Williams, C., & Rasmussen, C. (1995). Gaussian Processes for Regression. In *Advances in Neural Information Processing Systems* (Vol. 8). MIT Press.

Witczak, M. (2007). *Modelling and Estimation Strategies for Fault Diagnosis of Non-Linear Systems*. Berlin, Heidelberg: Springer. doi: 10.1007/978-3-540-71116-2_1

Automatic translation from System Dynamics to other formalisms with application to socio-bio-physical systems

John Tinnerholm[b,*] Mariano Zapatero[a] Adrian Pop[b] Peter Fritzson[b] Rodrigo Castro[a]

[a]*Universidad de Buenos Aires and ICC-CONICET, Argentina,* [b]*Linköping University, Linköping, Sweden*
[*]john.tinnerholm@liu.se

Abstract

System Dynamics is a modelling paradigm that has been applied to a wide range of systems, from economic to physical and from managerial to ecological. The main strength of the paradigm is its ease of use.
A System Dynamics modeller does not need to focus on equations; instead, models are expressed in terms of stocks and flows. Modelica, on the other hand, is an equation-based modelling language capable of multi-domain modelling using equations. It gives the user more freedom but requires more mathematical focus and skills.
Therefore, a unification of equation-based modelling and the System Dynamics paradigm is seen as highly beneficial. Advantages include the ability for System Dynamics modellers to use the tools available in the Modelica ecosystem. Furthermore, it allows the integration of System Dynamics models into Modelica models.
To achieve this goal, we developed an XMILE-to-Modelica translator that maps System Dynamics models represented in the XMILE standard exchange format to Modelica models. We also applied a Modelica-to-Julia translator to demonstrate the generality of the approach.
We translated several models to test the correctness of the translator. In particular, the Earth System Climate Interpretable Model (ESCIMO) was translated from its original version in the Vensim toolkit into the OpenModelica toolkit, and a correct validation was obtained by comparing simulation results between simulators.
Our work improves tool interoperability and further demonstrates the feasibility of using Modelica as a unified, standard language to integrate models created using System Dynamics, including large and complex socio-bio-physical systems.

1 Introduction

Currently, there exists no single unified environment for Modeling and Simulation (M & S). Instead, modelers utilize several heterogeneous modeling environments and paradigms. Consequently, modeler know-how and expert knowledge encoded in specific models are not necessarily available in all environments. The resulting branching for modeling practice imply that unnecessary idealizations and simplifications are made when modelers integrate concepts outside their area of expertise in their models. While there exist frameworks and methodologies such as the FMI-standard[1] that allow modelers to utilize models developed in different tools by constructing *Functional Mockup Units* (FMUs), these FMUs can then be simulated together using co-simulation or importing using model exchange (Gomes et al., 2018). However, this introduces additional complexity into the resulting composite model and thus complicates the analysis of the model variables and equations. Furthermore, the modeler may need to maintain models across a set of heterogeneous tools.

In order to facilitate modeling knowledge sharing and unify tools across different ecosystems, we present our approach of integrating models from the modeling paradigm *System Dynamics* (SD) into both the Modelica and Julia ecosystems by means of automatic translation: from XMILE into Modelica, and then from Modelica into Julia demonstrating the usefulness of Modelica as a unified equation-based language. To this end, we translated a rather complex SD model, the climate model *ESCIMO* (Earth System Climate Interpretable Model) (Randers et al., 2016) to test the aforementioned mapping. Furthermore, our approach also includes:

- The possibility to export SD models encoded in XMILE to FMUs.
- Integration of preexisting Modelica models into SD models.
- Increase model exchange between the Modelica, Julia, and SD communities.
- Increased tool support for SD models including bifurcation analysis and visualization.

[1] Accessed 2023-05-02: fmi-standard.org

- Interactive scripting of SD models in the Julia and Modelica environments.

1.1 Organization

The remainder of this paper is organized as follows: We introduce System Dynamics, Modelica and Julia in Section 2, Section 3, and Section 4 respectively. Following these sections, an extended motivation behind our work is presented in Section 6. The climate model, ESCIMO used as the motivating example for this work is presented in Section 5, the implementation of the translator from XMILE to Modelica is presented in Section 7, and the simulation results are presented in Section 8. Finally, we outline directions for future research in Section 9.

2 System Dynamics

System Dynamics (SD) is a modeling paradigm developed by the group of Jay Forrester at MIT in the 1950s (Forrester, 2007). In SD, modelers develop their models as webs of interacting positive and negative feedback loops, using the notion of stocks and flows as building blocks. Stocks represent the accumulation of an inventory (such as fish in a lake or cars at a dealership). Flows represent rates of change to such inventories. Using these notations SD modelers can develop models for complex systems, such as climate models or large socioeconomic models, a well-known model being the World3 global model (D. L. Meadows et al., 1974). Several environments support SD modeling, including *Stella*[2], *Vensim*[3], and *Simantics System Dynamics*[4].
To increase operability between different tools in the System Dynamics community, OASIS developed the *Interchange Language for System Dynamics* (XMILE) standard[5]. XMILE is an XML-based open exchange format that encodes System Dynamic models.

3 Modelica

Modelica is an equation-based object-oriented acausal modeling language developed by the Modelica Association. Modelica aims to be a unified language for equation-based modeling of (but not limited to) cyber-physical systems (Fritzson, 2015). Several toolkits support the Modelica language, including: *Dymola*[6] by Dassault Systèmes, *Modelon Impact* by Modelon[7], and the OpenModelica environment (Fritzson et al., 2020) by the Open Source Modelica Consortium (OSMC).

[2]Accessed 2023-05-06 Stella
[3]Accessed 2023-05-02: Vensim
[4]Accessed 2023-05-06: sysdyn.simantics.org
[5]Accessed 2023-05-06: XMILE
[6]Accessed 2023-05-02: Dymola
[7]Accessed 2023-05-02: Modelon Impact

Modelica differs from the SD paradigm because it supports both causal and acausal modeling and also object-orientation. As a consequence, SD models may be expressed in Modelica; however, not all Modelica models can easily be expressed using classic SD notation since Modelica is a more universally applicable formalism.
Due to the universal application of the equation notation of Modelica, there are libraries that can be used for the development of SD models with Modelica, one of the most well-known being the *System Dynamics* visual library (Cellier, 2008). Also, Modelica tools can be used for simulation as the backend of other tools such as the Simantics System Dynamic Tools (Lempinen et al., 2011). In this case, the models are expressed using the SD formalism, and they are automatically translated internally to Modelica to be simulated using the OpenModelica framework.

4 Julia

Julia is a programming language developed with a strong focus on numerical computing along with powerful metaprogramming capabilities (Bezanson et al., 2017). In recent years, the Julia language has received increased attention being awarded the Wilkinson price for Numerical Software in 2019. Due to this focus, several M&S environments have been developed for the Julia Language, with *ModelingToolkit.jl* (MTK) (Ma et al., 2021) being one of the most well known. To combine the power of Modelica and Julia, we have previously developed a framework capable of translating Modelica models into Julia (Tinnerholm et al., 2022).

5 The Earth System Climate Interpretable Model (ESCIMO)

ESCIMO (Randers et al., 2016) is an SD model that represents the global climate system, focusing on a time range from 1850 to 2100, and including factors such as fluctuations in sea levels and global temperature.
In an article published in Nature in 2020, the model was extended to simulate the global climate up to the year 2500. The model's forecast predicts that even if man-made greenhouse gas emissions were to stop in 2020, the global temperature would still continue to rise (Randers & Goluke, 2020).
The main components of the ESCIMO climate model and their interactions are depicted in Figure 1. The ESCIMO model consists of 1181 equations and variables.

5.1 ESCIMO and Earth3

ESCIMO has previously been integrated as a submodel in a larger socio-bio-physical model named

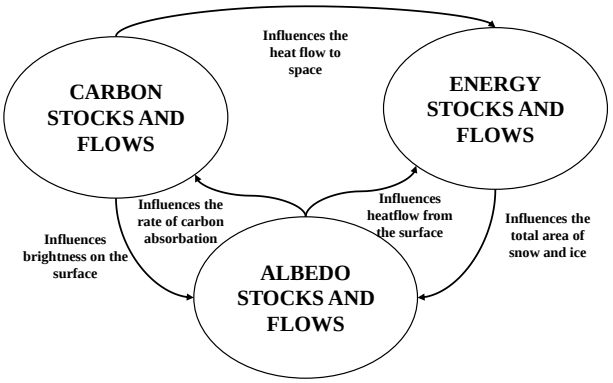


Figure 1. The three sectors of the ESCIMO climate model as described by (Randers et al., 2016).

Earth3.

The *Earth3* model was developed to examine the effect on the planetary boundaries if 14 out of the total 17 global Sustainable Development Goals (SDGs) agreed by the UN in 2015 were to be fulfilled (Randers et al., 2019). The conclusion of the simulation experiments are that it is not possible for humanity to achieve these 14 SDGs while at the same time not violating the planetary boundaries by 2030 or 2050 if the business-as-usual scenario (as defined by the author) continues.

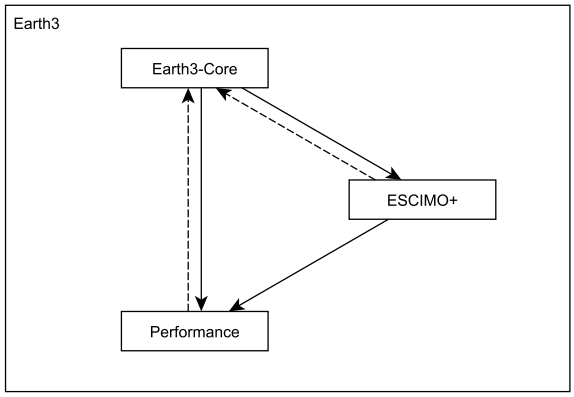


Figure 2. High-level overview of the Earth3 model where a variant of the ESCIMO mode called ESCIMO+ and Earth3-Core are sub-models. The dashed lines illustrate possible future feedback loops.

A high-level overview of the model is depicted in Figure 2. The model consists of three key sub-models:

- The ESCIMO model that models the global climate.
- The Earth3-Core model that models the socio-economic behavior.
- The performance model that calculates the performance with respect to the SDGs.

The Earth3-Core model (Randers et al., 2019) was developed in Microsoft Excel as a spreadsheet model, whereas ESCIMO+ was developed using the SD paradigm, as previously discussed. In its current formulation, without flows closing loops from ESCIMO+ back to Earth3-Core, a changing global climate will not affect the behavior of humanity as represented by the Earth3-Core model. Still, Randers et al. emphasize that the lack of these feedback loops has a greater effect on model variables after the year 2050, which was beyond the duration of the simulation experiment presented in the paper. This fact serves as a motivation behind the work presented in this article and, as explained in the introduction, translators from one modeling paradigm to another can yield substantial benefits.

6 Motivation

There exists a plethora of heterogeneous modeling and simulation tools. Although solutions exist that allow modelers to integrate models from different tools, such as the FMI standard, not all tools support this. In other cases, such as for the Earth3 model in Section 5, this is achieved by using the integration capabilities of existing tools; however, as discussed, this also imposes different limitations for each case. Simulating models using co-simulation adds extra complexity such as the selection of suitable master algorithms and a suitable step size, and might require mastery of several modeling paradigms as well as domain-specific tools in order to develop and maintain several models in tandem. Also, different scientific disciplines are accustomed to using different tools and languages to express their models; this leads to an ecosystem of modeling techniques that can undermine the development of more complex systems. Hence, in this work we propose to use Modelica as a unified, formal, and standard language to integrate models created both in spreadsheets and System Dynamics (SD).

6.1 Why Modelica

We argue that Modelica is a good fit for a unified language given it is open, standardized, object-oriented, and equation-based. As a consequence, it supports both acausal and causal modeling. This allows causal models encoded for instance, in SD to be encoded in Modelica. An example of an SD model is given in Figure 3.

A Modelica model for the SD model in Figure 3 is available in Listing 1. This exemplifies how Modelica's inheritance and composition permit an advantageous component-based approach to reduce duplicated equations in our model.

While there exists research proposing a similar component-based approach for the SD paradigm

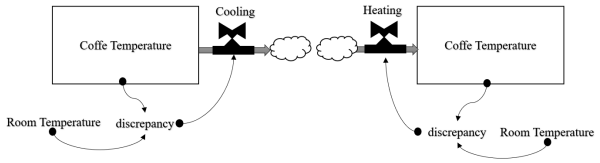


Figure 3. A simple SD model modeling heat over time in a coffee cup. This example is adapted from (D. H. Meadows, 2008).

(Bauer & Bodendorf, 2006) it has yet to reach mainstream adoption (Yeager et al., 2014).
Similarly to many SD tools, Modelica environments usually include a graphical notation that modelers can use to compose models using drag and drop. Examples of graphical modeling libraries for SD in the context of Modelica include the *System Dynamics* library (Cellier, 2008).

```
model CoffeeCup
  parameter Real cStartTemp = 60;
  parameter Real roomTemp = 18;
  Real roomTemperature(start = roomTemp);
  Real coffeeTemperature(start = cStartTemp);
  Real tmpChange;
  Real discrepancy;
equation
  roomTemperature = 18;
  tmpChange = discrepancy / 10;
  der(coffeeTemperature) = -tmpChange;
  discrepancy = coffeeTemperature - roomTemperature;
end CoffeeCup;

model Scenarios
  CoffeeCup coffeeCupHeating(cStartTemp = 5);
  CoffeeCup coffeeCupCooling(cStartTemp = 60);
  Real hotCoffeeCooling = coffeeCupCooling.tmpChange;
  Real hotCoffeeHeating = coffeeCupHeating.tmpChange;
end Scenarios;
```

Listing 1. Modelica model of the SD coffee cup model depicted in Figure 3. Here we use inheritance via modification to enable the two scenarios.

6.2 Why Julia

In addition to Modelica, we successfully experimented with translating the resulting Modelica version of the ESCIMO model to Julia.
There are several reasons for this translation. It exemplifies the ease of translation from a standard modeling language to other languages, and it provides access to the simulation runtime of OpenModelica.jl. The latter comes with extensions to the Modelica language for so-called *Variable-Structure-Systems* which allows conditional changing equations models during simulation. Hence, models simulated in this environment can be further modified to include scenarios where the dynamics of models radically change during the course of a simulation (Tinnerholm et al., 2022).

Access to the Julia ecosystem also comes with several advantages such as a wide set of scientific machine learning tools enabling domain-aware and physics-informed learning, state-of-the-art tools for bifurcation analysis[8] and interactive visualization[9] to name a few.
To conclude, in this section we have provided an extended discussion to exemplify the advantages of an automatic translation from System Dynamics to other formalisms. For further details, we refer to (Castro, 2019) which provides an extended discussion of this topic in the context of global models.

7 Mapping XMILE To Modelica

XMILE is a standard format that allows the interchange of SD models between toolkits. In order to map XMILE to Modelica an initial proof-of-concept translator was written in Python. The translator works by mapping entities described in the XMILE standard to corresponding entities in Modelica. For brevity, we will not describe all elements of the translator here (the full source code of the translator and the resulting models are available upon request).

```
<model>
  <sim_specs> <!-- OPTIONAL-->
  ...
  </sim_specs>
  <behavior> <!-- OPTIONAL-->
    ...
  </behavior>
  <variables> <!-- REQUIRED -->
    ...
  </variables>
  <views> <!-- OPTIONAL-->
   ...
  </views>
</model>
```

Listing 2. High level structure of an SD model encoded in XMILE (OASIS, 2015)

Listing 2 describes the overall structure of the model tag in XMILE. The current translator from XMILE to Modelica enumerates all variables and all equations of an XMILE document. Then, for each variable tag, it maps it one-to-one into a Modelica variable while keeping auxiliary information (such as units and dimensions). Currently, the translator supports the following categories of variables:

- **auxiliary**
- **stock**
- **flow**
- **delay1i**
- **delay3i**

[8] Accessed 2023-05-09 BifurcationKit.jl
[9] Accessed 2023-05-09 Interact.jl

- **delay3**
- **smooth3**
- **smooth**
- **sample_if_true**

Each category for each variable is saved both to be encoded in the final Modelica model and to generate the correct equations. Likewise, the initial values of each variable are used to construct the initial equations of the resulting Modelica model.
The equations of the model are constructed based on the category of each variable. The equations for auxiliary variables are translated *verbatim* since they may be mapped to simple algebraic equations. However, other categories of variables need to be transformed into an equivalent Modelica construct. The XMILE standard (OASIS, 2015) defines the stock as:

$$stock_t = stock_{t-dt} + dt \cdot (inflows_{t-dt} - outflows_{t-dt})$$

While this form is suitable for explicit solvers typically used in SD environments, in Modelica, the time step is not available directly during the simulation, so instead, this is reformulated as

$$der(stock) = inflows - outflows$$

where $der(stock)$ is the continuous time derivative.
The mapping for a subset of these categories to the corresponding Modelica equation is presented in Table 1.

Table 1. Subset of Modelica to SD type matchings

SD Type	Modelica Formulation
stock	$der(stock) = inflows - outflows$
smooth	$der(smooth) = \frac{input - smooth}{averagingTimeVariable}$
flow	$flow = inflow$

We tested the translator with the ESCIMO model described in Section 5.
Since ESCIMO is a part of Earth3 and takes some input from spreadsheets, we also needed to integrate an Excel parser in the translator.
The components of the spreadsheet model as defined in Excel were mapped to Modelica lookup tables[10]
Once this mapping was complete, we validated the model by running it using OpenModelica. The translation to Julia was simple, as described in Section 4; an existing compiler from Julia to Modelica was used for this purpose. The Julia compiler was validated by comparing the simulation results of the resulting simulation code with that obtained from OpenModelica. An excerpt of the ESCIMO model translated to Modelica is available in Listing 3. The full Modelica and the resulting full Julia models are available upon request.
This section has presented the XMILE to Modelica translator capable of translating a significant subset of XMILE as used by the ESCIMO model to Modelica. The next part of this paper will present the results of our validation experiments.

[10] In Julia, these tables were defined and implemented using a foreign function interface, OMRuntimeExternalC.jl, accessed 2023-05-16.

```
model ESCIMO
  constant Real Future_volcanic_emissions(unit =
  ↪  "GtVAe/yr") = 0.0 "CONST";
  constant Real Albedo_Antarctic_sens(unit = "fraction") =
  ↪  0.7 "CONST";
  constant Real
  ↪  Annual_pct_increase_CH4_emissions_from_2015_pct_yr(unit
  ↪  = "1/yr") = 0.0 "CONST";
...
initial equation
  Antarctic_ice_volume_km3 =
  ↪  Antarctic_ice_volume_in_1850_km3   "STOCK";
  Arctic_ice__on_sea__area_km2 =
  ↪  Arctic_ice_area_in_1850_km2   "STOCK";
  C_in_permafrost_in_form_of_CH4 = 1200.0   "STOCK";
...
equation
...
  der(DESERT_Mkm2) =
  ↪  flow_Shifting_GRASS_to_DESERT_Mkm2_yr -
  ↪  flow_Sifting_DESERT_to_GRASS_Mkm2_yr   "STOCK";
  der(Fossil_fuel_reserves_in_ground_GtC) =   -
  ↪  flow_Man_made_fossil_C_emissions_GtC_yr   "STOCK";
  der(GRASS_area_burnt_Mkm2) =  flow_GRASS_burning_Mkm2_yr
  ↪  - flow_GRASS_regrowing_after_being_burnt_Mkm2_yr
  ↪  "STOCK";
...
  UNIT_conversion_for_CH4_from_CO2e_to_C =  1/(16/12 *
  ↪  Global_Warming_Potential_CH4)   "AUX";
  UNIT_conversion_for_CO2_from_CO2e_to_C =  12/44   "AUX";
  UNIT_conversion_from_MtCH4_to_GtC =  1 /( 1000 / 12 *
  ↪  16)   "AUX";
...
  flow_SW_surface_absorption=SW_surface_absorption
  ↪  "FLOW";
  flow_GRASS_runoff=GRASS_runoff  "FLOW";
  flow_NATURE_CCS_Fig3_GtC_yr=NATURE_CCS_Fig3_GtC_yr
  ↪  "FLOW";
...
end ESCIMO
```

Listing 3. Excerpt from the translated ESCIMO model, showing initial equations and equations for some of the models' stocks and flows.

8 Simulation Results

During the course of our work, we experimented with several iterations of the ESCIMO model. We generated a corresponding Modelica model for the three ESCIMO models presented in (Randers et al., 2016, 2019; Randers & Goluke, 2020). This section, however, will present our validation of the ESCIMO model as presented in (Randers & Goluke, 2020). It should be noted that there are several configurations (or scenarios) in which one can simulate the model. The validation presented in this section concerns Scenario 1. The results were gathered from three simu-

lations: the publicly available ESCIMO model (baseline), the Modelica translation of this model, and the Julia model.
In our experiments, we compared the simulated values at every decade between 1850 to 2500 using the translated models and the reference model for the following variables:

- *Temperature surface anomaly compared to 1850* (Celsius), that is, the difference in average global surface temperature compared to 1850.
- *pH in warm surface water*, that is, the acidity of warm surface water.
- *CO2 Concentration in PPM*, that is, the concentration of carbon dioxide in the atmosphere.

The following numerical solvers were used:

- Runge-Kutta-4[11] for the Vensim simulation.
- DASSL for the OpenModelica[12] simulation with variable step-size, the absolute and relative tolerance was set to $1E-6$.
- Rodas5[13] for the Julia simulation with variable step-size, The absolute and relative tolerance was set to $1E-6$.

The resulting plots of these variables for the ESCIMO model are depicted in Figure 4. From the plots in the figure we observe that there is no visible difference between the three models. The graphs reveal that the translated models generate the same outcomes with minimal variations.
The percentage difference between the original model and the translated Modelica model for the **Temperature Anomaly** is presented in Figure 5. The graph shows that the difference between the original model and the resulting Modelica variable was, at most around 1%.
A plot highlighting the difference between the Julia and Modelica model for the variable **Temperature Anomaly** can be seen in Figure 6. As in the previous experiment, the difference between the variables was far below one percent, so it is not shown. In Figure 6, we can see that there are no significant differences between the Julia and OpenModelica environments. The largest difference in values observed between the Modelica and Julia models in Figure 6 occurred at $t = 1970$. Here, the value was ≈ 0.0552 and ≈ 0.0571 for the Modelica and Julia models, respectively, a difference around 3.5%. To compare, the value reported by the Vensim simulation was ≈ 0.0547. Hence, the difference between the Julia and Vensim model was

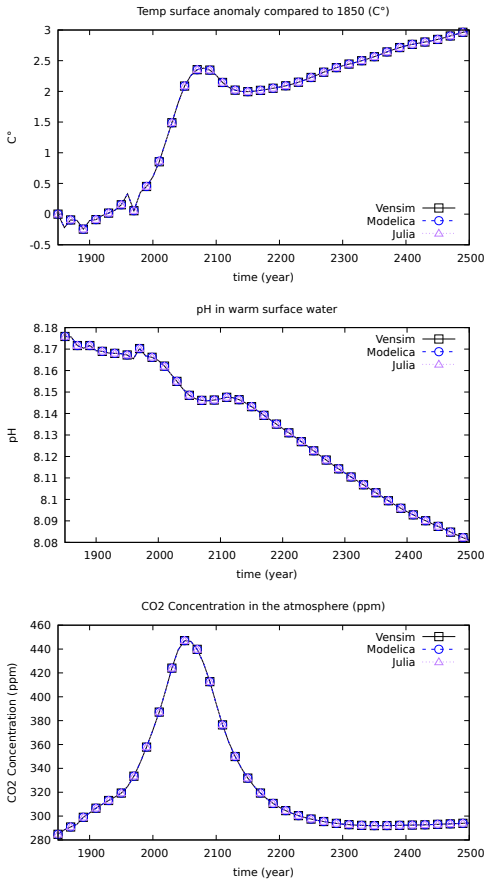


Figure 4. Graphs showing the simulation result of different variables from the year 1850 to the year 2500 for the Vensim, Modelica and Julia.

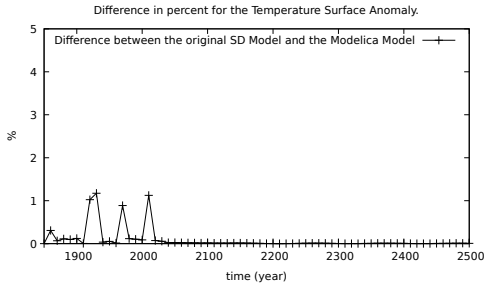


Figure 5. Difference in percent between the original SD Model simulated in Vensim and the Modelica model produced by the translator.

$\approx 4.29\%$. The reason for this divergence is due to how the Julia Simulation Engine handles a series of hybrid discrete events that occur in 1970. To summarize, the experiment shows a very small divergence from the original model; furthermore, we can observe that the dynamics of the resulting equations are the same. These differences in values between the Modelica and Vensim SD Model are due to differences in the numerical solvers used in the experiment.

[11] integration.html Accessed 2023-08-21
[12] OpenModelicaUsersGuide/latest/solving.html Accessed 2023-08-21
[13] DiffEqDocs/stable/solvers/ode_solve/ Accessed 2023-08-21

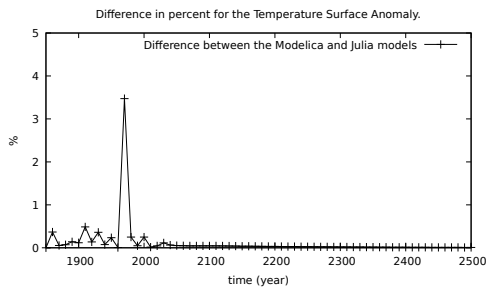


Figure 6. One excerpt for our comparison experiment that shows the difference in percent for the temperature surface anomaly variable between the Modelica and Julia models for each decade between 1850 and 2500.

9 Conclusion and Future Work

Simulation-based assessment of socio-bio-physical systems necessarily involves a wide range of knowledge domains. Different scientific disciplines tend to use different tools and languages to express their models. This leads to an ecosystem of modeling techniques that can undermine the development of more complex systems. In this work, we have presented a translator capable of translating models from one formalism, System Dynamics, to an object-oriented equation-based formalism as defined by the Modelica language. By so doing, we gained the ability to extract models from any tool that can export System Dynamics to XMILE and create a correct and equivalent Modelica and Julia model using a new automatic parser and translator. The existence of standard intermediate representation formats such as XMILE has been fundamental in achieving this goal.

We validated our efforts by comparing the fidelity of the translated model to the original Vensim model. In our experiments, it was shown that the difference between the Vensim and Modelica models was negligible, at most around 1% for the variables that were compared, see Figure 5. The overall dynamics of the translated system remained the same see the graphs in Figure 4. Hence, we can draw the same conclusions as (Randers & Goluke, 2020). As the main goal was to investigate the fidelity of the translated models to the original model, certain aspects, such as accuracy options of the numerical solvers and computational time, were not investigated in detail. However, both the OpenModelica environment and the Julia environment support a wide variety of industrial strength solvers capable of simulating models with more than tens of thousands of equations and variables under controlled accuracy (Ma et al., 2021; Fritzson et al., 2020; Rackauckas & Nie, 2017).

As future work, it remains to increase the number of functions and blocks supported by the translator between XMILE to Modelica, since for the present project the scope was set at what is necessary to meet the needs of the ESCIMO model. Further, it could be interesting to investigate the dynamics of a complete Earth3 model described in Section 5 by replacing the spreadsheet model with a Modelica model and activating the now-disabled feedback loops.

Also, as the model is now available in Julia and Modelica, it would be interesting to examine insights that can be obtained using various powerful tools in the Julia framework, such as scientific machine learning. It could also be interesting to augment the ESCIMO model using the structural variability of equations present in (Tinnerholm et al., 2022) and with other complex models available in the wider Modelica ecosystem. Robust optimization-driven sensitivity analysis could be performed for Earth3 using the OMSens plugin available for the OpenModelica toolkit (Danós et al., 2017).

We hope our work on unifying heterogeneous modeling paradigms will increase interdisciplinary collaboration in science and industry and enable a wider community to gain additional insights into a system as complex as planet Earth.

Acknowledgement

This work has been supported by the Swedish Government in the ELLIIT project and has been partially funded by Vinnova in the context of the ITEA project EMBRACE. It has also been partially funded by the Grant UBACyT 2020 Modalidad II 20620190200002BA from the University of Buenos Aires. Furthermore, the authors gratefully acknowledge the financial support of the French Transmission System Operator RTE to the Open Source Modelica Consortium, which made this initial development possible. Furthermore, we wish to thank the authors of the ESCIMO model, Jørgen Randers, and Ulrich Goluke for providing data from the ESCIMO Model and assisting with our validation effort.

We also acknowledge the comments from the authors[14] concerning the acronym ESCIMO model: ”Please note that the acronym for the software model described in the ESD paper is now recognized to be culturally insensitive and inappropriate. The editors of the journal ESD, the journal owner European Geosciences Union, and the publisher Copernicus Publications foster equality, diversity, and inclusiveness in scientific exchange and do not condone in any way racism, discrimination, or cultural appropriation. The authors did not intend to insult any ethnic groups by using the acronym for this software model.”

[14] Accessed 2023-05-11: http://www.2052.info/escimo/

References

Bauer, C., & Bodendorf, F. (2006). Enhancing system dynamics modeling using a component-based approach. *International Journal of Simulation: Systems, Science & Technology*, *7*(6).

Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM review*, *59*(1), 65–98.

Castro, R. (2019). Open research problems: Systems dynamics, complex systems. In B. P. Zeigler, A. Muzy, & E. Kofman (Eds.), *Theory of modeling and simulation* (Third Edition ed., p. 641-658). Academic Press. doi: 10.1016/B978-0-12-813370-5.00035-3

Cellier, F. E. (2008). World3 in modelica: creating system dynamics models in the modelica framework. In *Proceedings of the 6th international modelica conference* (Vol. 2, pp. 393–400).

Danós, A., Braun, W., Fritzson, P., Pop, A., Scolnik, H., & Castro, R. (2017, dec). Towards an OpenModelica-based sensitivity analysis platform including optimization-driven strategies. In *Proceedings of the 8th international workshop on equation-based object-oriented modeling languages and tools* (Vol. 17, pp. 87–93). New York, NY, USA: ACM. doi: 10.1145/3158191.3158206

Forrester, J. W. (2007). System dynamics—the next fifty years. *System Dynamics Review: The Journal of the System Dynamics Society*, *23*(2-3), 359–370.

Fritzson, P. (2015). *Principles of object-oriented modeling and simulation with modelica 3.3: A cyber-physical approach* (2nd ed.). Wiley-IEEE Press.

Fritzson, P., Pop, A., Abdelhak, K., Ashgar, A., Bachmann, B., Braun, W., … Östlund, P. (2020). The OpenModelica Integrated Environment for Modeling, Simulation, and Model-Based Development. *Modeling, Identification and Control*, *41*(4), 241–295. doi: 10.4173/mic.2020.4.1

Gomes, C., Thule, C., Broman, D., Larsen, P. G., & Vangheluwe, H. (2018, may). Co-simulation: A survey. *ACM Comput. Surv.*, *51*(3). doi: 10.1145/3179993

Lempinen, T., Ruutu, S., Karhela, T., & Ylén, P. (2011). Open source system dynamics with simantics and openmodelica. *Proceedings of the International System Dynamics Society*.

Ma, Y., Gowda, S., Anantharaman, R., Laughman, C., Shah, V., & Rackauckas, C. (2021). *Modelingtoolkit: A composable graph transformation system for equation-based modeling.*

Meadows, D. H. (2008). *Thinking in systems: A primer*. Chelsea Green Publishing.

Meadows, D. L., Behrens, W. W., Meadows, D. H., Naill, R. F., Randers, J., & Zahn, E. (1974). *Dynamics of growth in a finite world*. Wright-Allen Press Cambridge, MA.

OASIS. (2015). *Xml interchange language for system dynamics (xmile) version 1.0.* Retrieved from `http://docs.oasis-open.org/xmile/xmile/v1.0/os/xmile-v1.0-os.pdf`

Rackauckas, C., & Nie, Q. (2017). Differentialequations.jl – a performant and feature-rich ecosystem for solving differential equations in julia. *The Journal of Open Research Software*, *5*(1). doi: 10.5334/jors.151

Randers, J., & Goluke, U. (2020). An earth system model shows self-sustained thawing of permafrost even if all man-made ghg emissions stop in 2020. *Scientific reports*, *10*(1), 18456.

Randers, J., Golüke, U., Wenstøp, F., & Wenstøp, S. (2016). A user-friendly earth system model of low complexity: The escimo system dynamics model of global warming towards 2100. *Earth System Dynamics*, *7*(4), 831–850.

Randers, J., Rockström, J., Stoknes, P.-E., Goluke, U., Collste, D., Cornell, S. E., & Donges, J. (2019). Achieving the 17 sustainable development goals within 9 planetary boundaries. *Global Sustainability*, *2*, e24.

Tinnerholm, J., Pop, A., & Sjölund, M. (2022). A modular, extensible, and modelica-standard-compliant openmodelica compiler framework in julia supporting structural variability. *Electronics*, *11*(11). doi: 10.3390/electronics11111772

Yeager, L., Fiddaman, T., & Peterson, D. (2014). Entity-based system dynamics. In *Proceedings of the international system dynamics conference, delft.*

Data-driven reinforcement learning-based parametrization of a thermal model in induction traction motors

Anas Fattouh[a]*, Smruti Sahoo[b]

[a] *School of Innovation, Design and Technology (IDT), Mäalardalen University, Eskilstuna, Sweden,*

[b] *Alstom, Västerås, Sweden*

*Corresponding author: anas.fattouh@mdu.se

Abstract

Monitoring the temperature of induction traction motors is crucial for the safe and efficient operation of railway propulsion systems. Several thermal models were developed to capture the thermal behaviour of the induction motors. With proper calibrating of the thermal model parameters, they can be used to predict the motor's temperature. Moreover, calibrated thermal models can be used in simulation to evaluate the motor's performance under different operating conditions and find the optimal control strategies.

Parameterization of the thermal model is usually performed in dedicated labs where the induction motor is operated under predefined operating conditions and calibrating algorithms are then used to find the model's parameters. With the development of digital tools, including smart sensors, Internet of Things (IoT) devices, software applications, and various data collection platforms, operational data can be collected and used later to calibrate the parameters of the thermal model. Nevertheless, calibrating the model's parameters from operational data collected from different driving cycles is challenging as the model has to capture the thermal behaviour from all driving cycles' data.

In this paper, a data-driven reinforcement learning-based parametrization method is proposed to calibrate a thermal model in induction traction motors. First, the thermal behaviour of the induction motor is modelled as a thermal equivalent network. Second, a reinforcement learning (RL) agent is designed and trained to calibrate the model parameters using the data collected from multiple driving cycles. The proposed method is validated by numerical simulation results. The results showed that the trained RL agent came up with a policy that adeptly handles diverse driving cycles with different performance characteristics.

Keywords: Railway, Propulsion system, Traction motor, Induction motor, Thermal model, Parametrization, Data-driven, Reinforcement learning, Calibrating, Optimization

1. Introduction

Traction motors are subjected to varying operating and environmental conditions due to the dynamic loads over the operation cycle. The transient loads may cause overloading of the drive components which causes extra heat load. Operations causing overheating of the motor parts are of significant concern as they may lead to stator winding failure and accelerated ageing. Furthermore, to be able to exploit the motor's maximum utilization, it is essential that its operation is optimized to make it cost-effective.

On the other hand, induction motors (IMs) are the most used motors in railway propulsion applications to date because of their mechanical robustness and high overload capabilities. The added advantages are their low cost and the possibility of employing multiple drives connected to a single converter (Nategh *et al.*, 2020). However, their performance varies nonlinearly with temperature, frequency, saturation, and operating point which makes temperature monitoring essential for the safe and reliable operation of the motor.

The thermal limits of these motors are associated with the winding insulation material which is classified based on its temperature withstanding capacity. There are several established direct or indirect means for estimating the temperature in motor parts. Direct methods such as installing contact-based sensors in the stator, and rotor are the simplest means for measurement. However, the data transmission in the rotating parts has to be carried out with the help of end slip rings, or telemetry means. Regardless, installing sensors requires integration effort and additional cost and adds complexity due to their inaccessibility for replacement in case of failures or detuning. Hence model-based measurement techniques have been rather focused in the past decade (Ramakrishnan *et al*., 2009; Wilson, 2010). Here the temperatures can be estimated from the temperature dependent electrical parameters both off-line and online

manner. In these approaches, it is imperative that the model dynamic behaviours are accurately accounted for to avoid any estimation errors. These are also invasive in nature and create disturbance to the normal operation.

Computational fluid dynamics (CFD) and heat equation-based finite element analysis (FEA) are powerful techniques for accurate temperature estimation. However, due to their rigorous modelling effort and high computational power and time, they have been excluded from real-time monitoring upfront (Kirchgässner *et al*., 2021).

An alternative and computationally light temperature estimation technique is using the lumped-parameter thermal network (LPTN) model. An LPTN model summarizes the heat transfer process and can be represented in thermal equivalent circuit diagrams with knowledge of the used material and geometry and based on heat transfer theory. In this model, the thermal parameters are thermal resistances and capacitance values (Chen *et al*., 2020). The LPTNs can be designed to estimate the temperatures at several or fewer locations in the motor. However, the LPTN model needs accurate distributed motor losses, surface contact thermal conductance and heat transfer convection coefficients information, hence, needs to be calibrated empirically (Huber *et al*., 2014).

The complexity of LPTN models depends on the number of chosen nodes in the network. Generally, white box LPTN, based on pure analytical equations, are more accurate but are endowed with many thermal parameters, which could be complex to calculate in practice. A low-order structure with fewer nodes is computationally lightweight. In the past, several reduced order models categorized as light grey (5-15 nodes), and dark grey (2-5 nodes) LPTN models were structured to estimate the temperature with good accuracy. In this approach, only the dominant heat transfer pathways are represented, hence, expert domain knowledge is essential for the correct choice of not only their parameter values but also for their structural design (Wallscheid and Böcker, 2015). These models have proven to have good estimation performance, provided the parameters are identified appropriately. The use of a reduced order LPTN method would require estimations of many parameters that are not well known or possible to calculate using analytical equations. Thus, the identification of the parameters is an important step in these studies.

To date, several research works on the parameterization of the LPTN model have been performed using empirical measurements by applying different methods (Guemo *et al.*, 2013; Huber *et al.*, 2014; Xiao and Griffo, 2020). The proposed identification procedure varies the parameter values until the used LPTN model gives the same results as the experimental ones. As described by (Wallscheid, 2021), parameter identification can be made in a local approach or a global approach. In the work presented by (Huber *et al.*, 2014), a three-node LPTN is parametrized based on a global approach. A sequence of interdependent identification steps was followed, and the experimental data are used to find the thermal parameters. The model uses the measurement-based loss inputs available with motor electronic control unit (ECU) quantities, such as motor speed and electric currents. The parameter identification approach has been built on the idea of mapping the linear time-varying parameters to a set of time-invariant models operating within a certain chosen environment. Thus, a consistent parameter set for the whole operating region could be obtained with the adaptation of the relevant boundary conditions through various identification cycles. While the global approach is more robust in capturing all operating regions of the motor than the local approach, it can also be problematic if the parameter landscape to be identified is large and highly nonlinear in nature. Hence it is complex to find the parameter values near the global parameter optimum (Wallscheid, 2021). However, the global approach captures the nonlinearity in the form of a parametrizable function and, hence, they are potentially more versatile compared to the global and local approaches.

The use of inverse methods is also popular for finding thermal parameters. In the work presented by (Guemo *et al.*, 2013), the identification of the parameters is made by solving the optimization problem using determinist inverse-based methods such as the Gauss-Newton method, the Levenberg-Marquardt method, and stochastic inverse-based methods such as the Genetic Algorithms. These concept methods are used to minimize the residuals between measured and calculated temperatures. (Sciascera *et al.*, 2017) employed a tuning procedure based on a sequential quadratic programming iterative method for obtaining the uncertain thermal parameters of the thermal network. However, the computation cost of such tuning procedures is high due to the time-variant nature of the parameters. Furthermore, to improve computational efficiency, the dependence of the state matrix on the phase current is approximated with polynomial approximation.

The temperature rise in the electric motor occurs due to the electro-mechanical power conversion losses. Winding and core losses and mechanical and windage losses are the prominent ones for an induction motor. The winding losses can be calculated for a given winding resistance and measured current. However, the winding resistance changes based on the temperature which is a state variable in the thermal matrix. Furthermore, the core

loss is not measurable. A usual approach to determining iron losses is measuring total power losses and subtracting winding losses. Hence, all errors in the determination of total and winding losses directly add up to an error in the iron losses values. To deal with these uncertainties, (Gedlu *et al.*, 2020) used an extended iron loss model as input to a low-order LPTN model for temperature estimation. The loss inputs as a form of spatial loss model calculate individual core losses for each node. In addition to the heat transfer coefficients, the uncertain parameters in the core loss equations are calibrated in their possible searching space using particle swarm optimization (PSO) to minimize the estimation error in comparison to empirical measurements.

Xiao and Griffo (Xiao and Griffo, 2020) in their work presented an online measurement-informed thermal parameter estimation using a recursive Kalman filter method. While a Pulse-Width Modulation (PWM)-based estimation method is utilized for rotor temperature measurement, the temperatures in three nodes such as stator core, winding, and rotor are predicted. The input losses for the LPTN model are derived based on a model-based approach and with the use of Finite Elements (FE) analysis. The identification problem is formulated as a state observer with eight states. Three of the states correspond to the nodes' temperatures and the rest five states represent the unknown thermal resistances parameters in the LPT network. The non-linearity of the model is dealt with continuous updated linearization the extended Kalman filter method.

The growing interest and upsurge in machine learning (ML) techniques in the past decade make these potentially viable tools in the area of automated monitoring and motor drive control. A pure ML model, i.e. a model without expert knowledge of any classic fundamental heat theory, can be trained to estimate the temperature empirically. In this case, the model parameters are fitted based on collected testbench/observation data only (Kirchgässner *et al.*, 2021; Wallscheid, 2017). The widely used ML algorithm is the linear regression technique which has low computational complexity and is used for temperature predictions (Kirchgässner *et al.*, 2019; Zhu, 2019). However, as linear regression is a linear time-invariant, it does not capture the dynamics of the motor model.

In the field of sequence learning tasks and in high dynamics, recurrent neural networks and convolutional neural networks are the state of the art in classification and estimation performance. In the study conducted by (Kirchgässner *et al.*, 2019), deep recurrent and convolutional neural networks with residual connections are empirically evaluated for predicting temperature profile in the stator teeth, winding, and yoke as well. The concept is to parameterize neural networks entirely on empirical data, without being driven by the domain expertise. Furthermore, supervised ML models are also investigated for online parameter estimation such as rotor resistance and mutual inductance in the control system of an induction motor (Wlas *et al.*, 2008). In the presented work, a simple two-layer artificial neural network (ANN), consisting of an input layer, one hidden layer, and an output layer, is trained by minimizing the error between the rotor flux linkages based on an induction motor analytical voltage model and the output of the ANN-trained model. Feedforward and recurrent networks are used to develop an ANN as a memory for remembering the estimated parameters and for computing the electrical parameters during the transient state.

While pure data-driven ML models[1] are effective in predicting the temperature, their parameters are not interpretable and could not be designed with the low amount of model parameters as in an LPTN model at equal estimation accuracy. As a further development, to make expert knowledge-based calibration less desirable and to account for the uncertainties regarding the input power losses, (Kirchgässner W. W., 2023) proposed a deep learning-based temperature model where a thermal neural network is introduced, which unifies both consolidated knowledge in the form of heat-transfer-based LPTNs, and data-driven nonlinear function approximation with supervised machine learning.

The reinforcement learning (RL)-based methods are other promising data-driven techniques explored in the field of control of electric motor drives (Book *et al.*, 2021). RL methods enable learning in a trial-and-error manner and avoid supervision of each data sample. The algorithm requires a reward function to receive the reward signals throughout the learning process. Thus, the control policy could be improved on a continuous basis based on the measurement feedback (Sutton and Barto, 2018).

In this paper, a data-driven reinforcement learning-based parametrization method is proposed to calibrate a thermal model of an induction traction motor.

The rest of the paper is organized as follows. Section 2 presents the parametrization of an LPTN Model for an induction motor. Section 3 explains the developed RL framework to calibrate the parameters of the parametrized LPTN model. The dataset and the training process are given in Section 4. Section 5 shows and discusses the results of the calibrated thermal model. Concluding remarks and future work are given in Section 6.

[1] Models without expert knowledge.

2. Parametrizing the Thermal Model

From a thermal point of view, the motor is modelled with four nodes: stator winding (node 1), stator core (node 2), rotor winding (node 3) and rotor core (node 4). The thermal equivalent network is illustrated in Figure 1 with thermal capacitances, to which a power (heat) source is connected, and with thermal conductance among the nodes and to the cooling air.

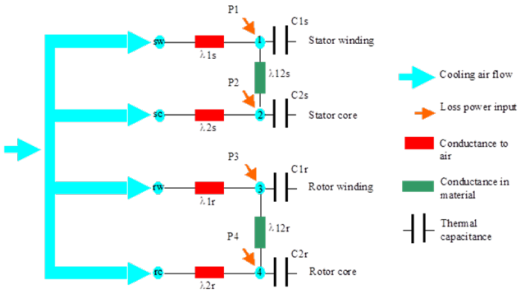


Figure 1: Lumped Parameter Thermal Network Model

Thermal capacitances C_{1s}, C_{2s}, C_{1r}, and C_{2r} values are calculated analytically from the geometry and material information of the motor. The capacitance for stator yoke C_{1s} is the sum of the capacitance of stator housing, stator back iron, stator tooth and flange mounted. The stator winding capacitance C_{2s} includes the capacitance for the stator winding and the end winding capacitances. The capacitance for stator yoke C_{1r} is the sum of the capacitance of rotor yoke, and rotor bars. The rotor winding capacitance C_{2r} includes the capacitance for the rotor winding and the end winding capacitances.

The thermal conductance λ_{1s}, λ_{2s}, λ_{1r}, λ_{2r} vary with the airflow due to the convection. The model shown in Figure 1 can be represented mathematically by the following first-order differential system:

$$P_1 = C_{1s}\frac{dT_1}{dt} + \lambda_{1s}(T_1 - T_{env}) + \lambda_{12s}(T_1 - T_2) \quad (1)$$

$$P_2 = C_{2s}\frac{dT_2}{dt} + \lambda_{2s}(T_2 - T_{env}) + \lambda_{12s}(T_2 - T_1) \quad (2)$$

$$P_3 = C_{1r}\frac{dT_3}{dt} + \lambda_{1r}(T_3 - T_{env}) + \lambda_{12r}(T_3 - T_4) \quad (3)$$

$$P_4 = C_{2r}\frac{dT_4}{dt} + \lambda_{2r}(T_4 - T_{env}) + \lambda_{12r}(T_4 - T_3) \quad (4)$$

where T_i is the temperature at the corresponding node i. The temperatures of the cooling air at the four nodes (marked as sw, sc, rw and rc in Figure 1) are assigned to the environment (or ambient) temperature T_{env}.

The losses at the four nodes in Figure 1 are distributed as shown in Table 1 and they can be calculated as follows:

$$P_1 = K_{stemp}P_{cu1} + K_{stray}P_{stray} + K_{harm}P_{harm} \quad (5)$$

$$P_2 = K_{Pfe}P_{fe} \quad (6)$$

$$P_3 = K_{rtemp}P_{cu2} + (1 - K_{stray})P_{stray} + (1 - K_{harm})P_{harm} \quad (7)$$

$$P_4 = (1 - K_{Pfe})\ P_{fe} \quad (8)$$

where $P_{cu1}, P_{cu2}, P_{stray}, P_{harm}, P_{fe}$ are the stator copper loss, rotor copper loss, stray loss, harmonic loss, and iron loss respectively. The coefficients K_{stemp}, K_{rtemp}, K_{stray}, K_{harm} and K_{Pfe} are the corresponding losses coefficients.

Table 1: Loss Distribution in the LPTN Model.

Node	Winding Losses	Stray Losses	Harmonic Losses	Iron Losses
1	x	x	x	
2				x
3	x	x	x	
4				x

The losses in Equations (5)-(8) can be calculated as follows (Kral *et al.*, 2013; Filizadeh, 2013; Maroteaux, 2016; Nasir, 2020; IEC/TS, 2010):

$$P_{cu1} = R_1\ I_1^2 \quad (9)$$

$$P_{cu2} = R_{21}\ I_{21}^2 \quad (10)$$

$$P_{stray} = P_{SUP}\left(\frac{f}{f_{nom}}\right)^{1.5}\left(\frac{I_1}{I_{1,nom}}\right)^2 \quad (11)$$

$$P_{fe} = K_f\, f^{\alpha} B_{max}^{\beta} \quad (12)$$

where I_1, I_{21} are the stator and rotor currents respectively, R_1 and R_2 are the stator and rotor winding resistances respectively which depend on the temperature according to following equations:

$$R_1 = R_{1,20} * (1 + \alpha_{R1} * (T_1 - 20)) \quad (13)$$

$$R_{21} = R_{21,20} * (1 + \alpha_{R2} * (T_3 - 20)) \quad (14)$$

where $R_{1,20}$, $R_{21,20}$ are the stator and rotor winding resistance at 20 °C and α_{R1}, α_{R21} the temperature coefficient of stator and rotor respectively. In Equation (11), f is the stator frequency with a nominal value f_{nom}, I_1 is the stator current with a nominal value $I_{1,nom}$ and P_{SUP} is equivalent rated input power. In Equation (12), K_f is a constant that depends on the material properties and the core geometry, f is the frequency of the magnetic field, B_{max} is the peak magnetic flux density in the core and α and β are empirically determined constants. The harmonic losses P_{harmo} is measured at few operation points and included as a look- up table in the loss model.

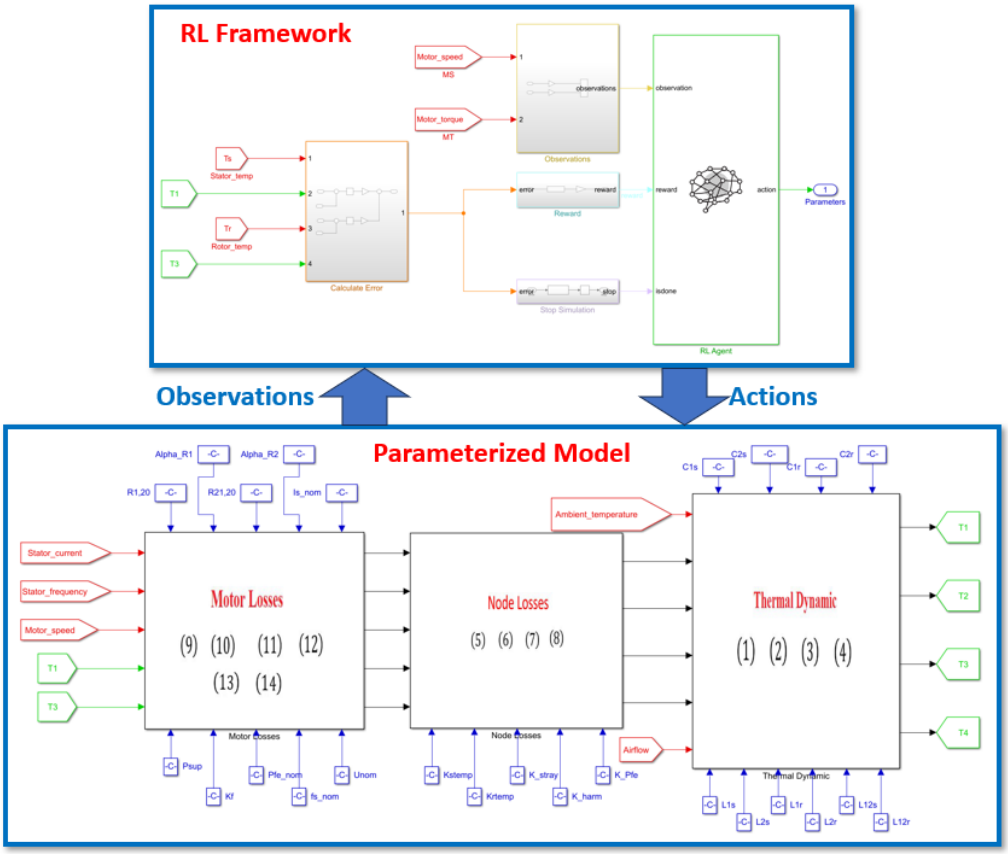


Figure 2: Reinforcement Learning Framework

3. The Reinforcement Learning Framework

The model developed in the previous subsection has been implemented as a parameterized model and employed within a reinforcement learning (RL) framework as shown in Figure 2.

3.1. The Parameterized Model

The parameterized model represents the dynamic evolution of the temperature in the induction motor according to the model developed in Section 2. The model inputs (the red signals) are the stator current, the stator frequency, the motor speed, the airflow of the cooling air and the ambient temperature. The model outputs (the green signals) are the temperatures (T_1, T_2, T_3, T_4) at the four nodes (P_1, P_2, P_3, P_4) respectively of the LPTN model shown in Figure 1. The blue signals in Figure 2 represent the parameters of Equations (1)-(14) explained in Section 2.

3.2. The RL Framework

The RL framework consists of the RL agent, the observations, the reward function and the actions.

3.2.1. The Observations

The observations represent the information that the RL agent can sense from the environment (the parameterized model) during the operation of the induction motor. Some observations, such as the motor speed (MS) and the motor torque (MT), are used directly by the agent to sense any changes in the motor's operating conditions. Other observations, such as the measured stator and rotor temperatures (Ts, Tr) and the model outputs (T_1, T_3), are used to calculate the reward. It should be noted that T_2 and T_4 could be considered among the observations if there are related measurements. However, in this work, there are no measurements related to T_2 and T_4.

3.2.2. The Reward Function

The reward function produces a value that reflects the effectiveness of the agent's actions in the environment. This value serves as a critical signal guiding the agent's learning process towards achieving its goals effectively. The reward value encapsulates the objectives of the RL problem, which is, in our context, minimizing the error between the measured temperatures (Ts, Tr) and the model's output temperatures (T_1, T_3). Hence, the reward function is given by:

$$r = \frac{-\omega_1}{\omega_2|T_s - T_1| + \omega_3|T_r - T_3| + \omega_4} \qquad (15)$$

where ω_1, ω_2, ω_3 and ω_4 are positive weights.

3.2.3. The Actions

The actions represent all the possible values of the model parameters (the blue signals in the

parameterized model). Actions are computed by the agent based on the observations and the reward value and using a policy that is optimized during the training process to maximize the expected cumulative reward over time. The policy is essentially the agent's strategy for selecting actions in different situations to achieve its objectives efficiently.

3.2.4. The RL Agent

The RL agent is composed of two main elements: a policy and a learning algorithm. The policy maps the observations with actions to be taken while the learning algorithm updates the policy parameters, based on the actions, observations and rewards, to maximize the expected cumulative long-term reward. During the learning process and depending on the learning algorithm, the agent retains two types of models: critic and actor models. The critic model predicts the expected cumulative reward (Q-Value) from a given observation and action that is later used by the actor model to return the action that maximizes the predicted discounted cumulative long-term reward (Sutton and Barto, 2018).

In this work, a TD3[2] algorithm is used in the RL framework shown in Figure 2. TD3 agent works in a continuous environment and has improved policy model performance over time (Dankwa *et al.*, 2019; Nicola and Nicola, 2021). Moreover, the episodic training paradigm enables the TD3 agent to select different training datasets after each episode. This will allow the agent to find the optimal policy (the thermal model parameters) from measurements recorded during different driving cycles, which is the main objective of this work.

In the following section, the dataset and the training process are explained.

4. Training the RL Agent

4.1. The Dataset

The dataset represents the data recorded from the induction motor during the operation of nine different driving cycles. It is composed of all the red signals shown in Figure 2, i.e., the motor speed, the motor torque, airflow, the stator current and frequency, the motor voltage, the stator winding temperature and the rotor winding temperature.

It should be noted that the data are recorded at different sampling frequencies with some missing values that require resampling the dataset and interpolating the missing values.

It should also be noted that the dataset is not used directly by the RL framework to train the agent, but it is used in the parameterized model which is considered an unknown environment to the agent.

4.2. The TD3 Structure

The TD3 algorithm employs one actor model (network) and two critic models (networks) as shown in Figure 3.

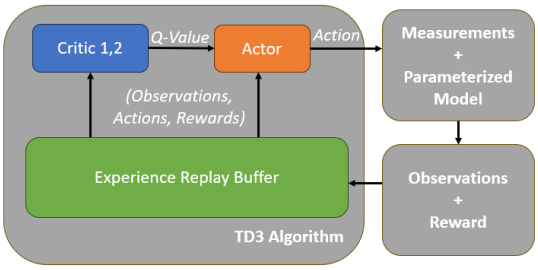


Figure 3: Simplified TD3 Structure

The actor and critic models are approximation models, such as a deep neural network, that are trained from *(Observations, Actions, Rewards)* sampled from the experience replay buffer.

4.3. The Training Process

The RL agent is trained following an episodic scheme. Episodes are used to model the concept of a task or problem that the agent is trying to learn (Sutton and Barto, 2018). An episode refers to a single run of the agent's interaction with the environment, starting from an initial state and continuing until a terminal state is reached.

Each episode consists of a specific number of discrete time steps. At each discrete time step:

1. The actor applies an action based on the current observation and expected Q-value.
2. The new observation, action and reward are stored in the experience replay buffer.
3. A random batch of experiences is sampled from the experience replay buffer and used to update the parameters of the critic models by minimizing a loss function across all sampled experiences.
4. After some specified steps, the parameters of the actor model are updated using a sampled policy gradient that maximizes the expected discounted reward.

5. Results and Discussion

5.1. Preprocessing the Dataset

In this work, nine driving cycles have been used to represent the unknown environment which is used to train the agent to find the optimal values of the thermal conductance λ_{1s}, λ_{2s}, λ_{1r} and λ_{2r}. As mentioned previously, data are recorded at different sampling frequencies with some missing values that

[2] Twin-Delayed Deep Deterministic Policy Gradient

require resampling the dataset and interpolating the missing values.
Figure 4 shows a sample of measured motor speed and motor torque from one driving cycle (DC1) after resampling and interpolating the missing values.

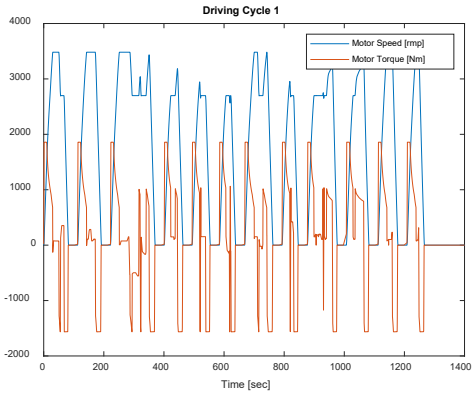


Figure 4: A Sample of Motor Speed and Torque (DC1)

5.2. Training the RL Agent

The RL agent was trained using the actor and critic deep neural network (DNN) models with the layers depicted in Figures 5 and 6.

	Name	Type	Activations	Learnable Prope...
1	input_1 2 features	Feature Input	2(C) × 1(B)	-
2	fc_1 256 fully connected layer	Fully Connected	256(C) × 1(B)	Weights 256 × 2 Bias 256 × 1
3	relu_body ReLU	ReLU	256(C) × 1(B)	-
4	fc_body 256 fully connected layer	Fully Connected	256(C) × 1(B)	Weights 256 × 256 Bias 256 × 1
5	body_output ReLU	ReLU	256(C) × 1(B)	-
6	output 5 fully connected layer	Fully Connected	5(C) × 1(B)	Weights 5 × 256 Bias 5 × 1
7	tanh Hyperbolic tangent	Tanh	5(C) × 1(B)	-
8	scale Scaling layer	ScalingLayer	5(C) × 1(B)	-

Figure 5: Layers of the Actor DNN Model

	Name	Type	Activations	Learnable Prope...
1	input_1 2 features	Feature Input	2(C) × 1(B)	-
2	fc_1 256 fully connected layer	Fully Connected	256(C) × 1(B)	Weights 256 × 2 Bias 256 × 1
3	input_2 5 features	Feature Input	5(C) × 1(B)	-
4	fc_2 256 fully connected layer	Fully Connected	256(C) × 1(B)	Weights 256 × 5 Bias 256 × 1
5	concat Concatenation of 2 inputs along dimension 1	Concatenation	512(C) × 1(B)	-
6	relu_body ReLU	ReLU	512(C) × 1(B)	-
7	fc_body 256 fully connected layer	Fully Connected	256(C) × 1(B)	Weights 256 × 512 Bias 256 × 1
8	body_output ReLU	ReLU	256(C) × 1(B)	-
9	output 1 fully connected layer	Fully Connected	1(C) × 1(B)	Weights 1 × 256 Bias 1 × 1

Figure 6: Layers of the Critic 1 and 2 DNN Models

The training steps explained in Subsection 4.2 were applied with the parameters shown in Table 2.

Table 2: Training Parameters.

Property	Value
Max Episodes	1000
Max Steps per Episode	9990
Averaging Window Length	100
Stop Training Value	-190
Agent Sample Time	0.1

Figure 5 shows the training process where the blue curve represents the episode reward, the red curve represents the average reward and the orange curve represents the estimated cumulative rewards at the beginning of each training episode. The figure shows that the agent learned an optimal policy (parameters) after 101 episodes.

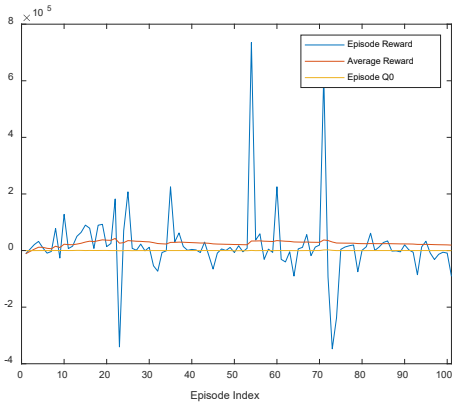


Figure 5: Training Process

5.3. Validating the Trained Agent

Validation of the RL agent usually involves periodically evaluating the learned policy directly in the environment.
When evaluating the learned policy with different driving cycles, the agent came up with the following parameters:
$\lambda_{1sn} = 1$, $\lambda_{1rn} = 40$, $\lambda_{2sn} = 20$, $\lambda_{2rn} = 0.0001$, $Q_n = 10^{-9}$
Figures 7 and 8 show the measured and model temperatures for driving cycle 1 (DC1) and driving cycle 8 (DC8) respectively.

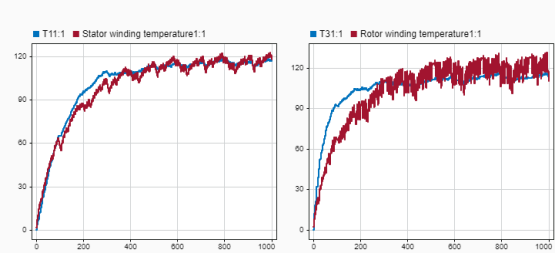


Figure 7: Model Parameters and Measured and Model Temperatures for DC1

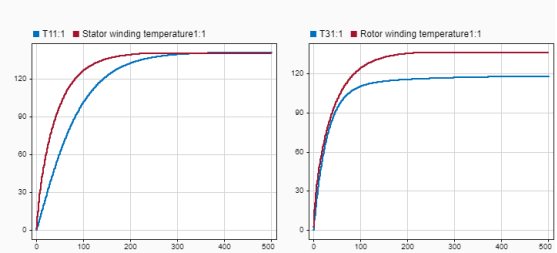


Figure 8: Model Parameters and Measured and Model Temperatures for DC8

6. Conclusion and Future Work

In the paper, a reinforcement learning framework is proposed for training an agent to find the parameters of the thermal model in induction traction motors. The framework has been applied to find the thermal conductance for the thermal network model from nine driving cycles.

By running different driving cycles, the trained agent came up with a policy that produces the parameters for the different driving cycles. The model with the calibrated parameters showed a good estimation of stator and rotor temperature.

In future work, other structures for the agent and the reward function will be considered to produce better temperature estimation.

Acknowledgment

This research work has received funding through the AIDOaRt project from the ECSEL Joint Undertaking (JU) under grant agreement No 101007350.

References

Book, G., Traue, A., Balakrishna, P., Brosch, A., Schenke, M., Hanke, S., Kirchgässner, W. and Wallscheid, O. (2021) 'Transferring online reinforcement learning for electric motor control from simulation to real-world experiments', *IEEE Open Journal of Power Electronics*, 2, pp. 187-201.

Chen, B., C. Wulff, K. Etzold, P. Manns, G. Birmes, J. Andert and S. Pischinger (2020) 'A comprehensive thermal model for system-level electric drivetrain simulation with respect to heat exchange between components', *19th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm), IEEE.*

Dankwa, S. and Zheng, W. (2019) 'Twin-delayed DDPG: A deep reinforcement learning technique to model a continuous movement of an intelligent robot agent', *Proceedings of the 3rd international conference on vision, image and signal processing*.

Filizadeh, S. (2013) *Electric machines and drives: principles, control, modeling, and simulation*. CRC press.

Gedlu, E. G., Wallscheid, O. and Böcker, J. (2020) 'Permanent magnet synchronous machine temperature estimation using low-order lumped-parameter thermal network with extended iron loss model', *The 10th International Conference on Power Electronics, Machines and Drives (PEMD 2020). IET.*

Guemo, G. G., Chantrenne, P. and Jac, J. (2013) 'Parameter identification of a lumped parameter thermal model for a permanent magnet synchronous machine', *2013 International Electric Machines & Drives Conference*. pp. 1316-132.

IEC/TS 60349-3 (2010) 'Electric traction - Rotating electrical machines for rail and road vehicles - Part 3: Determination of the total losses of converter-fed alternating current motors by summation of the component losses', https://webstore.iec.ch/publication/1830

Huber, T., W. Peters and J. Böcker (2014) 'Monitoring critical temperatures in permanent magnet synchronous motors using low-order thermal models', *2014 International Power Electronics Conference (IPEC-Hiroshima 2014-ECCE ASIA)*, IEEE.

Kirchgässner, W., Wallscheid, O. and Böcker, J. (2019) 'Deep residual convolutional and recurrent neural networks for temperature estimation in permanent magnet synchronous motors', *2019 IEEE International Electric Machines & Drives Conference (IEMDC)*. IEEE.

Kirchgässner, W., Wallscheid, O. and Böcker, J. (2021) 'Data-driven permanent magnet temperature estimation in synchronous motors with supervised machine learning: A benchmark', *IEEE Transactions on Energy Conversion*, 36(3), pp. 2059-2067.

Kirchgässner, W., Wallscheid, O. and Böcker, J. (2023) 'Thermal neural networks: lumped-parameter thermal modeling with state-space machine learning', *Engineering Applications of Artificial Intelligence*, 117, p. 105537.

Kral, C., Haumer, A. and Lee, S. B. (2013) 'A practical thermal model for the estimation of permanent magnet and stator winding temperatures', *IEEE Transactions on Power Electronics*, 29(1), pp. 455-464.

Maroteaux, A. (2016) *Study of analytical models for harmonic losses calculations in traction induction motors*. KTH, School of Electrical Engineering (EES).

Nasir, B. A. (2020) 'An Accurate Iron Core Loss Model in Equivalent Circuit of Induction Machines', *Journal of Energy*, 2020, pp. 1-10.

Nategh, S., A. Boglietti, Y. Liu, D. Barber, R. Brammer, D. Lindberg and O. Aglen (2020) 'A review on different aspects of traction motor design for railway applications', *IEEE Transactions on Industry Applications,* 56(3): 2148-2157.

Nicola, M. and Nicola, C.-I. (2021) 'Improvement of PMSM control using reinforcement learning deep deterministic policy gradient agent', *21st International Symposium on Power Electronics (Ee)*, pp. 1-6.

Ramakrishnan, R., Islam, R., Islam, M. and Sebastian, T. (2009) 'Real time estimation of parameters for controlling and monitoring permanent magnet synchronous motors', *2009 IEEE International Electric Machines and Drives Conference*. pp. 1194-1199.

Sciascera, C., Giangrande, P., Papini, L., Gerada, C. and Galea, M. (2017) 'Analytical thermal model for fast stator winding temperature prediction', *IEEE Transactions on Industrial Electronics*, 64(8), pp. 6116-6126.

Sutton, R. S. & Barto, A. G. (2018) *Reinforcement learning: An introduction*. MIT press.

Wallscheid, O. and J. Böcker (2015) 'Design and identification of a lumped-parameter thermal network for permanent magnet synchronous motors based on heat transfer theory and particle swarm optimisation', *17th European Conference on Power Electronics and Applications (EPE'15 ECCE-Europe)*, IEEE.

Wallscheid, O. (2021) 'Thermal monitoring of electric motors: State-of-the-art review and future challenges', *IEEE Open Journal of Industry Applications*, 2, pp. 204-223.

Wilson, S. D., Stewart, P. and Taylor, B. P. (2010) 'Methods of resistance estimation in permanent magnet synchronous motors for real-time thermal management', *IEEE Transactions on Energy Conversion*, 25(3), pp. 698-707.

Wlas, M., Krzeminski, Z. and Toliyat, H. A. (2008) 'Neural-network-based parameter estimations of induction motors', *IEEE Transactions on Industrial Electronics*, 55(4), pp. 1783-1794.

Xiao, S. and Griffo, A. (2020) 'Online thermal parameter identification for permanent magnet synchronous machines', *IET Electric Power Applications*, 14(12), pp. 2340-2347.

Zhu, Y., Xiao, M., Lu, K., Wu, Z. and Tao, B. (2019) 'A simplified thermal model and online temperature estimation method of permanent magnet synchronous motors', *Applied Sciences*, 9(15), p. 3158.

Simulation of distribution system for low temperature district heating in future urban areas – Case study of a planned city district in Gävle

Oskar Olsson[a,*] Mattias Gustafsson[b] Magnus Åberg[a]

[a]*Department of Civil and Industrial engineering, Uppsala University,* [b] *Faculty of Engineering and Sustainable Development, Gävle University*
[*]oskar.olsson.6790@student.uu.se

Abstract

In Europe, the prices of natural gas and electricity reached an all-time high in 2022. A way to mitigate high electricity costs is to expand district heating systems in urban areas, this will reduce electric load as well as increase the power generation possibilities in combined heat and power plants. District heating has been the dominant heat supply technology in urban areas in Sweden since the 1980s. However, as the energy efficiency in buildings increase, district heating distribution losses must be reduced to ensure a cost-efficient heat supply. This has led to the idea of the 4th-generation district heating which is characterized by low distribution temperatures. In this study, low-temperature district heating distribution in a planned future city district is simulated using a Python-based tool. Two different low-temperature distribution systems are investigated: 1) 2-pipe low-temperature system, and 2) a cascading 3-pipe low-temperature system. The focus is on simulating the distribution losses, temperature drop, and mass flow in the pipe network. The scope of the analysis also includes an investigation of the effect of lower return temperatures on the central district heating network. The results indicate that the low-temperature distribution system with the 2-pipe system performs better than the cascading system when considering distribution losses and temperature drop. The mass flow depends on the temperature demand in the heating systems in the buildings and is considerably high for both low-temperature distribution systems investigated.

1 Introduction

In Europe, record-high electricity and natural gas prices were noted in 2022. Several countries within the EU have introduced financial instruments and subsidies to hamper consequences of high energy prices (Eurostat, 2023). These are, however, short-term solutions that are not necessarily sustainable in the long run, this as the dependence on fossil fuel imports remains. A more long-term solution is to develop energy systems to be more efficient and thus reduce dependence on fossil fuels. A possible way to reduce the urban need for electricity and natural gas is to replace conventional gas- and electric heaters in buildings with district heating systems (DHS). A measure that would also enable increased co-production of electricity in combined heat and power plants and thereby yield an increased overall system efficiency (Colmenar-Santos et al., 2016).

DH development has traditionally aimed to reduce heat distribution temperatures in order to reduce distribution losses, which is likely to be of further importance as the share of new low-energy buildings increases meaning that distribution losses tend to increase relative the heat demand. The concept of "4^{th}-generation district heating" that was introduced by Lund et al. (Lund et al., 2014) is based on distribution temperatures below 70°C in order to deal with the challenge of a future high share of energy-efficient buildings.

The structures of DH networks in cities differs significantly depending on the local conditions. Access to industrial waste heat and the need for extra pumps as a result of large height differences in the system are two examples of aspects that influence the structure of a system. This means that when developing and expanding distribution networks, there is not one single solution that is optimal for all systems, instead individual solutions need to be developed and adapted for each system (Jakubek et al., 2023). System network models and simulations of temperatures, water flows and pressure drop, are thus potentially powerful tools to choose distribution techniques and design networks in order to achieve technically well-functioning and cost-efficient DH systems (Nguyen et al., 2020).

There are previous network-simulation studies that focus on the effects of reducing distribution losses by lowering the distribution temperatures. Pirouti et al. (Pirouti et al., 2013) have optimized flow and supply temperatures to minimize losses and total cost. Their results show that small pipe diameters, large pressure drops, and large differences in supply and return temperature in the system were advantageous. There are also studies focusing on developing simulation models for DH distribution. Valdimarsson (Valdimars-

son, 2012) and Press (Press, 2022) for instance, use graph theory to present distribution losses, temperature drops, pressure drops, and flows for each individual pipe in a system. Jakubek et al. (Jakubek et al., 2023) simulates the losses for different types of pipes. Also, studies exist concerning low temperature DH distribution but without network simulations. Werner for example, investigated to what extent different types of distribution techniques for low-temperature DH have been implemented in real systems (Werner, 2022), unfortunately the article does not include an analysis of the losses associated to different distribution techniques.

1.1 Aim of the project

The overall aim of this project is to simulate a low-temperature DHS in order to investigate the impact of different distribution technologies on heat losses, distribution temperatures, and mass flows. A case study is made for the planned city district called Näringen in the Swedish city of Gävle. The district contains a high share of energy-efficient buildings and therefore illustrates the challenges that DHS will face in the future. The city district is divided into 11 sub-areas that that will be sequentially built. One of these areas were chosen for the distribution network simulations. For this area, a DH network is designed and implemented in a simulation model. The question to be answered in this project is:

How do different distribution system configurations for low-temperature district heating in the investigated area differ regarding distribution losses, distribution temperatures, and mass flow?

2 Background

This section gives a brief introduction to the 4^{th}-generation DH concept and a description of the district Näringen, which is the case study object in this study.

2.1 4^{th}-generation district heating

The 4^{th}-generation DH has several similarities with, what is known as the third generation of DH. Pressurized water is the heat carrier and the pipes are prefabricated and in the ground. The main difference between the two generations is the lower supply temperatures for the 4^{th}-generation. This primarily motivated by reduced distribution losses that comes with reduced temperature difference between the ground and the pipes. However, this also means that waste heat from low-temperature sources such as data centres, can be used to a further extent in DHSs (Tofani, 2022). Efficient DH distribution is generally considered a prerequisite for the future whit higher building energy efficiency (Lund et al., 2014).

4^{th}-generation DH is, however, still in the development stage and is currently best suited for newly built residential areas as existing networks are designed and adapted to existing and less energy efficient building stocks. Several low-temperature systems have been built in Germany, but a few projects in Sweden have also been tested. In order to integrate new low-temperature systems in new energy-efficient districts with the existing high-temperature systems, the new system can be connected to the main DHS as a sub-system using heat exchangers. This to enable control of pressure and temperatures in the sub-system separate from the main system (Borglund, 2020).

2.2 Näringen

The district of Näringen is centrally located in the Swedish city of Gävle and has a total area of 232 hectares. Gävle municipality has an agreement with Region Gävleborg and the Swedish government to transform Näringen into a sustainable city district containing 6,000 homes. In return, the Swedish government is planning investments for infrastructure worth of 20 billion SEK in Gävle until 2040 (Gävle kommun, 2021). A map of the district and the planned sub-areas are presented in Fig. 1. This paper focus on sub-area 11 that contains 69 buildings, of which 53 are residential and 16 commercial buildings. There is no DHS in this area today, the designed pipe network presented here is therefore a possible design for a future system.

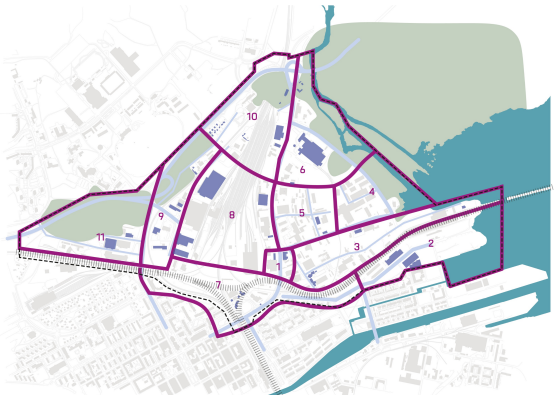


Figure 1. Overview of Näringen and the preliminary sub-area breakdown of the district.

3 Methodology

This section describes the process from designing the distribution systems in the sub-area and implementing it into the simulation tool.

3.1 Model description

The simulation tool used in the project is based on the Python programming language and created by Arvid Press (Press, 2022). The tool is further developed in

this project to increase the precision of the calculations for heat losses and temperature drop. The two main changes were first, that the heat loss calculations was extended from only considering one single pipe above ground to instead consider two separate pipes in the ground. The second main improvement is an implementation of a minimum flow requirement in the system. The latter to illustrate a hot water circulation loop used to avoid extremely low water flows when space heating demand is low. The simulation tool is based on graph theory, which means that the DH system is implemented by defining all branches and connection points as nodes/points and all DH pipes as arcs/edges, and that the system components and its connections are described as a complete graph. A system can thus be described by a matrix where the rows correspond to the number of connections and the columns corresponds to the number of pipes. The elements in the matrix define the direction of the edges, 1 represents a start node, -1 an end node, and 0 that there is no connection between the nodes. This means that each column in the matrix can only have two nonzero elements, 1 and -1 , because each edge (DH pipe) must start and end somewhere. The matrix is therefore specific for every system, two examples of matrices and for further description of the mathematics behind the simulation tool are described in (Valdimarsson, 2012) and (Press, 2022). An illustration of a principal system as described in the simulation tool is presented in Fig. 2.

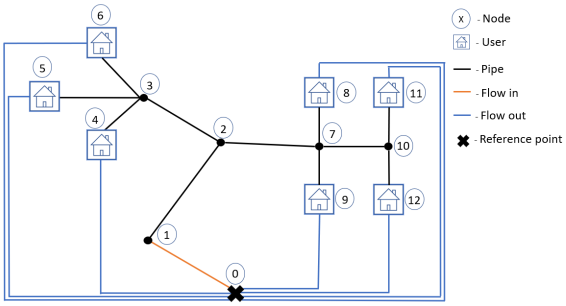


Figure 2. Example system illustrating the model principle.

3.2 Distribution network design

The distribution network in sub-area 11 is designed based on a map of the planned buildings and streets. The structure of the pipe network follows a tree-structure, thus there are no loops and only one inlet to the system. In addition to this, the inlet to the system is placed to limit the distance to the system's outermost connection point. The map was thereafter loaded into QGIS (GIS simulation program) and by using an integrated measuring tool, all pipe lengths could be determined and used to describe the system in the simulation tool.

One of the buildings in the area is excluded from the simulations since it is a parking garage and is assumed to not be heated. The layout of the distribution system and how it connects the buildings is illustrated in Fig. 3.

Figure 3. Distribution network in sub-area 11.

3.3 Distribution techniques

When distributing DH, the most commonly used technology is a 2-pipe system where the space heating and domestic hot water have the same supply and return pipes. This configuration works for both high and low distribution temperatures. In this project, a classic low-temperature system with a 60°C distribution temperature was simulated. The return temperature vary depending on the simulated building type. The supply temperature is restriced to a minimum of 60°C to avoid growth of legionella bacteria in the domestic hot water system (Fredriksen & Werner, 2014). A schematic view of the 2-pipe system can be seen in Fig. 4 where SH stands for "space heating" and DHW stands for "domestic hot water".

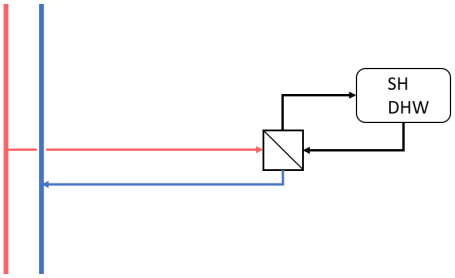


Figure 4. Pipe configuration 2-pipe system.

The second investigated distribution system configuration is a cascaded 3-pipe system, where the space heating supply pipe is connected to the main system's return pipe. The difference from from the above described 2-pipe system is that the space heating and domestic hot water supply is divided into two separate pipes. The 3-pipe configuration thus have the option of having different supply temperatures for the space heating and the domestic hot water, that the pipe for space heating can be disconnected during months without heat demand, and that each supply pipe can be

individually sized for the respective heat demand. In a low-temperature network with energy-efficient buildings, the 3-pipe system provides the possibility of lowering the temperature of the supply pipe for space heating below the legionella requirement, which potentially reduces distribution losses and enables the utilization of more low-temperature residual heat. The supply temperature for SH depends on the temperature requirement in each building type. A schematic view of the cascade-connected 3-pipe system can be seen in Fig. 5.

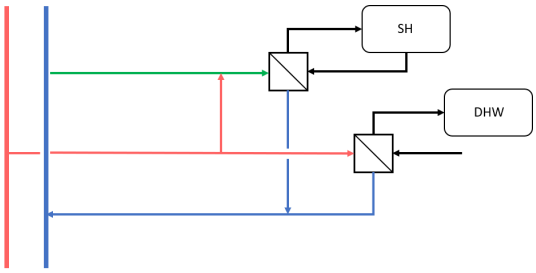


Figure 5. Pipe configuration of the cascade-connected 3-pipe system.

However, the 3-pipe systems need to have a back-up connection to the main systems supply pipe (or some other local high temperature source) to secure that the supply temperature during peak demand is sufficient and to ensure that the legionella requirement is met.

3.4 Simulations

This section describes the used formulas and parameters as well as how the simulations were executed.

3.4.1 Dimensioning

The first step was to size the pipes in the system (see Eq 1), which is done by calculating the system water flow demand using the design outdoor temperature (in Sweden known as DVUT) and the specific temperature demands of each building (Fredriksen & Werner, 2014). DVUT is defined as the average value of the outdoor temperature during the coldest day and is the temperature at which heating systems are dimensioned according to the Swedish national board of building, housing and planning building regulations (BBR) (Warfvinge & Dahlblom, 2010). Simulations are made with outdoor temperature data for 2022 provided by Gävle Energi AB, that owns the main district heating system in Gävle.

$$d_i = \left(\frac{8\lambda}{\pi^2 \rho P_{dl} \dot{m}_i^2} \right)^{\frac{1}{5}} \tag{1}$$

In Eq 1, P_{dl} is the pressure drop per meter and is assumed to be 200 pa/m, λ is the friction factor, ρ is the water density and $\dot{m}_i$ is the mass flow in each pipe. The calculations were done for both distribution system configurations and building types. The flow demand is used to calculate the pipe sizes and thereafter is the obtained pipe dimensions used to calculate the temperature drop in the system at DVUT. In addition to the pipe-sizing calculations, a simulation was also done over all hours of the year to calculate the total distribution losses. Calculations were made for both BBR buildings and passive houses, for detailed description of the building standards see (Israelsson, 2023).

3.4.2 Distribution losses

The distribution losses were simulated with hourly data using the equations for two insulated pipes in the ground, as described in chapter 5 pages 80-82 in (Fredriksen & Werner, 2014).

3.4.3 Cascaded system

In this part, calculations of return temperatures and flows for the cascaded 3-pipe system are presented. Mass flow and return temperature are calculated according to Eq. 2-5. T stands for temperature and $\dot{m}$ stands for mass flow. The indexing r,H represents return pipe of the main system, p,H represents the primary pipe of the main system and p,S represents the primary pipe in the secondary system. $\dot{m}_{tot,H}$ is the known total mass flow in the main system, see Tab. 1. The impact of the return temperature from the subsystem on the return temperature on the main system is calculated using a flow-weighted average value for the two temperatures.

$$T_{r,H}\dot{m}_{r,H} + T_{p,H}\dot{m}_{p,H} = T_{p,S}\dot{m}_S \tag{2}$$

$$\dot{m}_{r,H} = \dot{m}_S\left(1 - \frac{T_{p,S} - T_{r,H}}{T_{p,H} - T_{r,H}}\right) \tag{3}$$

$$\dot{m}_{p,H} = \dot{m}_S - \dot{m}_{r,H} \tag{4}$$

$$T_{ny,H} = \frac{(\dot{m}_{tot,H} - \dot{m}_{r,H})T_{r,H} + \dot{m}_S T_r}{\dot{m}_{tot,H}} \tag{5}$$

Eq. 3 determine the required flow from the main system's return pipe while required flow from the primary pipe is determined from Eq. 4. Eq. 5 calculates the new return temperature ($T_{ny,H}$) in the main system.

3.4.4 Mass flow

The mass flow at DVUT was simulated for the different building types. The required mass flow was compared to a traditional high-temperature system. Additionally, simulations for two different minimal flow levels a were performed to determine the impact of

the heat circulation-loops on the distribution temperatures. The minimum mass flow was initially and iteratively set to 0.02 kg/s to avoid significant losses. Thereafter, a minimum flow of 0.1 kg/s was investigated as it is suggested by Alros (Alros, 2015).

3.4.5 General input data

Constants used in the calculations are presented in Tab. 1. Due to a lack of measured data for the specific site, the ground temperatures for heat loss and temperature drop calculations are illustrated by mean values of the outdoor air temperature for different parts of the year; December for the pipe sizing calculations, annual mean for distribution loss calculations, and the summer months (June, July, and August) when investigating the minimum flow requirements (Fredriksen & Werner, 2014).

4 Results

This section presents the results from the simulations. In the first part, distribution temperature, return temperature, and distribution losses for the different pipe systems and building types are presented. In the second part, results showing the impact of a cascaded system on the main system temperatures is presented. Finally, the mass flow results are presented.

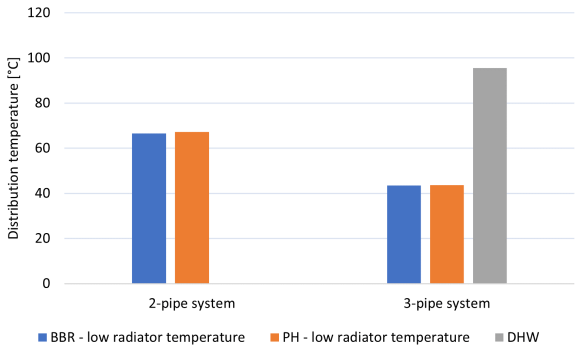
Table 1. Constant parameters for the simulations.

Parameter	Value	Unit
$T_{a,DVUT}$	-3.1	$°C$
$T_{a,dist}$	8.6	$°C$
$T_{a,summer}$	19.3	$°C$
$T_{p,H}$	99.7	$°C$
$T_{r,H}$	49.3	$°C$
$\dot{m}_{tot,H}$	15.8	kg/s

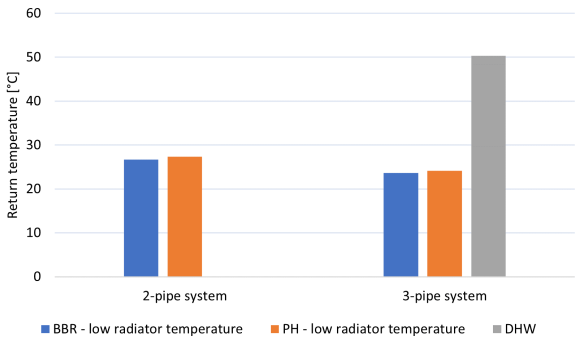
4.1 Temperatures and distribution losses

The temperatures and distribution losses for the subsystem are presented in Fig. 6. The temperature requirement in Fig. 6a and the return temperature in Fig. 6b are simulated at DVUT. It can be deduced from the figures that low radiator temperature for both BBR buildings and passive houses when using 2-pipe systems give similar results. Using a 3-pipe system, the

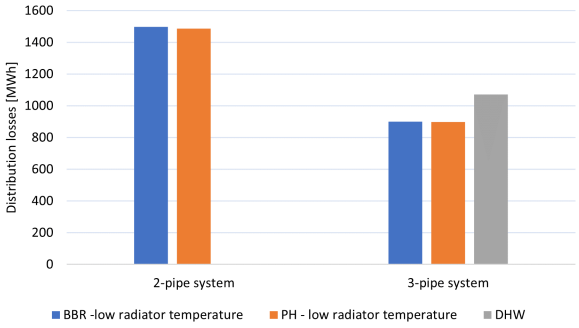
(a) Distribution temperature for 2- and 3-pipe systems.



(b) Return temperature for 2- och 3-pipe systems.



(c) Distribution losses for 2- och 3-pipe systems.



(d) Relative distribution losses for 2- och 3-pipe systems.

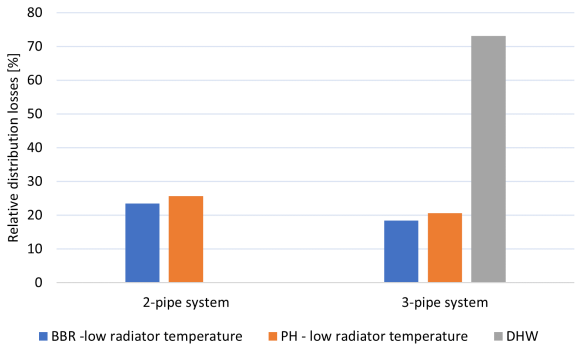


Figure 6. Distribution and return temperatures at DVUT and the annual absolute and relative distribution losses for buildings according to BBR and passive house standard in sub-area 11.

difference between BBR buildings and passive houses is small but there is a significant difference compared to the 2-pipe system. With a 3-pipe system, the temperature required for the domestic hot water is 96°C for the different building types since it only depends on the residents' hot water use (this is further explained in Section 5.1).

Fig. 6c and 6d shows absolute and relative distribution losses in the system. For the 2-pipe and 3-pipe systems, the difference in losses between the building types is small. The absolute distribution losses are lower for passive houses, while the relative distribution losses are lower for BBR buildings. The distribution losses for the domestic hot water account for a large proportion of the 3-pipe system and the total losses are greater than for the 2-pipe system for the different building types.

4.2 Cascaded system

The flows required from the main system pipes (return and primary) depending on the different building types, are presented in the two left stacks of bars in Fig 7. The right stack of bars presents the available flow in the main system pipes.

Fig. 7 shows that neither BBR buildings nor passive houses can be supplied with heat entirely from the main system's return pipe since the available mass flow ($15.8 kg/s$) is not sufficient. This means that a little flow from the primary pipe must be used, even though the temperature requirement (see Fig. 6a) is lower than the return temperature in the main system (see Tab. 1).

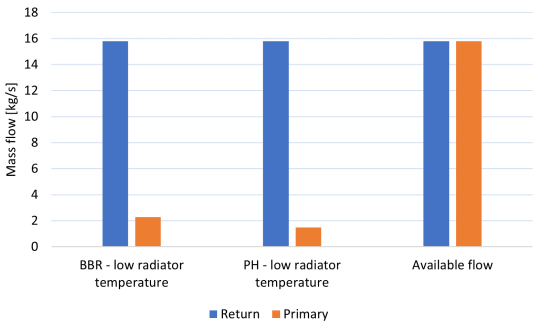


Figure 7. Mass flow requirements for the different building types in sub-area 11 when using a cascaded system.

The new return temperature for the sub system and the impact of the 3-pipe cascading system configuration on the return temperature in the main system is presented in Tab. 2. The results show that the greatest temperature reduction of the return in the main system is obtained with BBR buildings, which is explained by the higher flow requirements. For the sub-system, the new return temperature in the main system will be the same as the return temperature in the sub-system when the entire flow is used to supply the heat demand, see Fig. 6b.

Table 2. Temperature and temperature reduction in the main system return pipe for the different building types for sub-area 11.

	Return temperature [°C]	ΔT [°C]
BBR low temp. rad	23.6	25.7
PH low temp. rad	24.1	25.2

4.3 Mass flow

In this section, the results for the mass flow at DVUT for the low temperature systems compared to the mass flow at current standard radiator-temperatures levels are presented. The results from the analysis of the minimum flow requirements are also presented here.

4.3.1 Flow requirement at DVUT

Fig. 8 shows the flow demand for 2- and 3-pipe systems for the low-temperature radiators compared to current standard radiator-temperatures. The result shows that the flow requirement is lower with current temperature standards, when compared to the calculated temperatures for low-temperature radiators in (Israelsson, 2023). The largest flow is required for low-temperature radiators in BBR buildings, which is as expected due to the higher heat demand compared to passive houses.

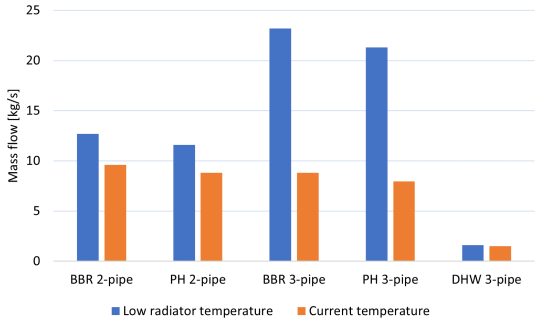


Figure 8. Flow demand with optimized temperature requirement and today's temperature for the different building types with 2- or 3-pipe distribution systems.

4.3.2 Sensitivity analysis of the minimum flow

When simulating the impact of the minimum flow on the temperature requirements and the total flow in the distribution system. This analysis is performed for the summer months when the heating demand is low and thus flow volumes in the system is generally small.

For higher heat demand periods the flow levels are significantly higher and heat circulation loops are thus not needed. The pipe sizes were obtained from the first part of the project when the system was sized for the heat demand at DVUT. As there is no need for space heating during summer, the 3-pipe system only uses the domestic hot water pipe in this case.

Tab. 3 presents the mass flow results that clearly shows that an increased minimum flow in the system reduces the temperature requirement in the subsystem, this goes for both 2- and 3-pipe systems. The mass flows, on the other hand, increases as a direct effect of the increased minimum flow.

Table 3. Distribution temperature and mass flow during the summer for sub-area 11 with different minimal flows.

	Dist. temp [°C]	Mass flow [kg/s]
2-pipe (0.02 kg/s)	105.8	1.6
2-pipe (0.1 kg/s)	67.8	6.8
3-pipe (0.02 kg/s)	84.6	1.6
3-pipe (0.1 kg/s)	64.5	6.8

5 Summary and Discussions

5.1 Results analysis

With the 3-pipe system, the temperature in the space heating pipe is lower when compared to the 2-pipe system. This, however, comes with low domestic hot water flows for the 3-pipe system (see Fig. 8), that cause large temperature drops in the system. This means that, in order to meet the temperature requirement of 60°C at the user, the temperature supplied to the systems needs to be as high as 90°C (see Fig. 6a). If the goal is to lower the temperature in the entire Gävle DH network and not only in Näringen, these results suggest that a 2-pipe system is better since the supply temperatures does not exceed 70°C for this system configuration.

The absolute distribution losses are marginally lower for passive houses compared to BBR buildings which is explained by the reduced heat demand in passive houses. The reason to why low-energy buildings with low radiator temperature requirements yield high distribution losses is because of relatively high flow requirements at DVUT (see Fig. 8). This high flow demand yield large pipe diameters, which means that the low flow volumes in summer leads low flow rates, high temperature drops, and in the end increased losses. The combined losses for space heating and domestic hot water for 3-pipe systems result in the total distribution losses being higher for both BBR buildings and passive houses. It is explained by the fact that the low flows for domestic hot water lead to large losses.

Analysis of the relative distribution losses in Fig. 6d shows that the lower absolute distribution losses for passive houses is combined with a relatively larger reduction of the heat demand, leading to increased relative losses for the system. Again, the problem with the domestic hot water supply in the 3-pipe systems is clear, the relative losses exceed 70% in sub-area 11, which in combination with the high-temperature requirements means that such a system design is not well suited for Näringen.

The results for the cascaded systems show that the mass flow in the main system return pipe is limited, but that it could supply at least parts of the heat demand in Näringen. A cascaded system in parts of Näringen may, however, cause problems in the future if the return temperature in the main system is reduced due to other efficiency measures. In that case, the potential for heat supply from the return pipe to Näringen would be further limited. Also, an investigation regarding the need for circulation pumps to maintain the pressure in the system is needed.

5.2 Summary

A model based on graph theory has been shown to be an effective tool when simulating DH networks. Designing and implementing future systems is possible if access to data on the heat demand, geographic location, and temperature requirements for the heating systems is available.

When using low-temperature DH for both space heating and domestic hot water, 2-pipe systems were shown to yield the lowest total distribution losses and distribution temperature requirements. 3-pipe systems have the potential to reduce distribution losses for space heating, but the low domestic hot water flows mean that the total losses and the temperature required in the system increase.

Finally, it is concluded that cascading entire or larger parts of Näringen to the return pipe of the nearby branch of the main network is not possible as the flow demand exceeds the available flow.

Acknowledgement

This work was supported by the Swedish Energy Agency, project Samspel mellan värme, kyla och elanvändning i ett bostadsområde med en hög andel solcellsproducerad el [grant number P2022-00442].

6 Nomenclature

DH	District heating
DHS	District heating system
SH	Space heating
DHW	Domestic hot water
BBR	Boverkets building regulations
PH	Passive house
DVUT	Design outdoor temperature

References

Alros, M. (2015). *Energikartläggning av vvc-systemet i flerbostadshus.* Retrieved from https://www.diva-portal.org/smash/get/diva2:808048/FULLTEXT01.pdf

Borglund, A.-S. (2020). *Framtidens fjärrvärme tar form.* Retrieved from https://www.energi.se/artiklar/framtidens-fjarrvarme-tar-form/ (2023-02-09)

Colmenar-Santos, A., Rosales-Asensio, E., Borge-Diez, D., & Blanes-Peiró, J.-J. (2016). District heating and cogeneration in the eu-28: Current situation, potential and proposed energy strategy for its generalisation. *Renewable and Sustainable Energy Reviews*, *62*, 621-639. Retrieved from https://www.sciencedirect.com/science/article/pii/S1364032116301149 doi: https://doi.org/10.1016/j.rser.2016.05.004

Eurostat. (2023). *Electricity production, consumption and market overview.* Retrieved from https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Electricity_production,_consumption_and_market_overview#Electricity_generation (2023-04-21)

Fredriksen, S., & Werner, S. (2014). *Fjärrvärme och fjärrkyla*. Lund: Studentlitteratur.

Gävle kommun. (2021). *Förstudie stadsomvandling näringen.* Retrieved from https://www.gavle.se/kommunens-service/bygga-trafik-och-miljo/planer-och-samhallsbyggnadsprojekt-i-gavle/pagaende-byggprojekt-i-gavle/naringen/ (2023-02-23)

Israelsson, K. (2023). *Värmebehov i byggnader i en planerad stadsdel med lågtempererad fjärrvärme som värmekälla.* Retrieved from https://uu.diva-portal.org/smash/get/diva2:1771089/FULLTEXT01.pdf

Jakubek, D., Oclon, P., Nowak-Oclon, M., Sulowicz, M., Varbanov, P. S., & Klemes, J. J. (2023). Mathematical modelling and model validation of the heat losses in district heating networks. *Energy*, *267*, 126460. Retrieved from https://www.sciencedirect.com/science/article/pii/S0360544222033461 doi: https://doi.org/10.1016/j.energy.2022.126460

Lund, H., Werner, S., Wiltshire, R., Svendsen, S., Thorsen, J. E., Hvelplund, F., & Mathiesen, V. (2014). 4th generation district heating (4gdh) integrating smart thermal grids into future sustainable energy systems. *Energy*, *68*, 1–11. doi: http://dx.doi.org/10.1016/j.energy.2014.02.089

Nguyen, T., Gustavsson, L., Dodoo, A., & Yao Ayikoe Tettey, U. (2020). Implications of supplying district heat to a new urban residentual area in sweden. *Energy*, *194*. doi: https://-doi.org/10.1016/j.energy.2019.116876

Pirouti, M., Bagdanavicius, A., Ekanayake, J., Wu, J., & Jenkins, N. (2013). Energy consumption and economic analyses of a district heating network. *Energy*, *57*, 149-159. Retrieved from https://www.sciencedirect.com/science/article/pii/S0360544213000984 doi: https://doi.org/10.1016/j.energy.2013.01.065

Press, A. (2022). *Utformning av modell för simulering av distributionsnät för fjärrvärme.* Retrieved from https://www.diva-portal.org/smash/get/diva2:1670033/FULLTEXT01.pdf

Tofani, A. (2022). *A case study on the integration of excess heat from data centres in the stockholm district heating system.* Retrieved from http://kth.diva-portal.org/smash/get/diva2:1723944/FULLTEXT01.pdf

Valdimarsson, P. (2012). *District heat distribution network.* Retrieved from https://orkustofnun.is/gogn/unu-gtp-sc/UNU-GTP-SC-18-27.pdf

Warfvinge, C., & Dahlblom, M. (2010). *Projektering av vvs-installationer*. Lund: Studentlitteratur.

Werner, S. (2022). Network configurations for implemented low-temperature district heating. *Energy*, *254*, 124091. Retrieved from https://www.sciencedirect.com/science/article/pii/S036054422200994X doi: https://doi.org/10.1016/j.energy.2022.124091

Models for Hydropower Plant: A review

Tajana Nepal[a,b,*], Diwakar Bista [b], Thomas Øyvang [a], Roshan Sharma [a]

[a] *University of South Eastern Norway,* [b] *Kathmandu University*
nepaltajana@gmail.com

Abstract

Hydro Power plant (HPP), being one of the most convenient options for power generation, has been modelled considering very wide aspects of their application. A model is simply a mathematical representation of a system and it may serve different purposes like dynamic simulation of hydro power, energy systems modelling involving policy making, condition monitoring, etc. The purpose of modelling HPPs may lead to various kind of models for a single Hydropower. This paper aims at reviewing hydropower models developed using different methods along with the purpose for modelling them. This will provide brief insights about state of the art on hydropower modelling and its emerging techniques. Furthermore, this paper presents in more detail about tracking the advancements in dynamic models for classical and variable speed hydropower plants highlighting the need for the development of more accurate models. The work mainly involves narrative review of published works on hydro power modelling techniques. Also, it includes systematic reviews about dynamic representation of hydropower plants. As this paper aims at presentation of hydro power models in a classified manner based on purpose of modelling, the areas of improvement in each type of model have been discussed. Models for control can be made to be more accurate by including more realistic featured like penstock dynamics, uncertainties, etc which further help in design of advanced control systems. There are several potential benefits of HPP modelling, such as optimizing plant performance, improving control, reducing maintenance costs, and enhancing overall system efficiency and reliability.

Nomenclature:

ANPC	Active Neutral Point Converter
CFD	Computational Fluid Dynamics
DFIC	Doubly Fed Induction Generators
FVM	Finite Volume Method
HVDC	High Voltage Direct Current
IGBT	Insulated Gate Bipolar Junction Transistor
MPC	Model Predictive Control
MMC	Multi-Level Converters
NPC	Neutral Point Converters
PMSG	Permanent Magnet Synchronous Generator
PSH	Pump Storage Hydropower
RLC	Resistance Inductance Capacitance
RTDR	Rotating Dynamic Response
SFR	Standstill Frequency Response

1 Introduction

Hydropower plants have been proven to be the most sustainable source of energy (Kumari Rupesh *et al.*, 2019; Shahgholian, 2020). Installed capacity of the hydropower all over the world comprises of approximately 20% of the world's electricity sources and 80% of the renewable sources (Shahgholian, 2020). This fact reflects the dire need to make the hydropower plants more efficient, more reliable and more economically viable. For this, more studies and research have to be performed, and more useful tools have to be developed. The primary step to be taken for this is to develop an appropriate model of hydropower systems which addresses the purpose of study.

Development of hydropower models has been carried out for different purpose and have undergone considerable improvement since the 90's. Accordingly, the methods used for modelling also vary to a wide category based on the application of the model. Some of them are developed for planning studies while others are developed for control, transient response, study of dynamics, condition monitoring, etc (de Mello *et al.*, 1992; Kishor, Saini and Singh, 2007; Valavi and Nysveen, 2018a; Liu *et al.*, 2019; Sapkota *et al.*, 2022). A single hydropower has many components starting from water reservoir and flow regions, mechanical rotating parts to the static electric parts in general(Quiroga OD, 2000; Rheinheimer *et al.*, 2023). Moreover, variable speed hydropower plants have converters as the additional elements and pump storage power plants have reversible pump-turbines (Nobile, Sari and Schwery, 2018a). Each component falling under the different disciplines of study aggregate to make

.

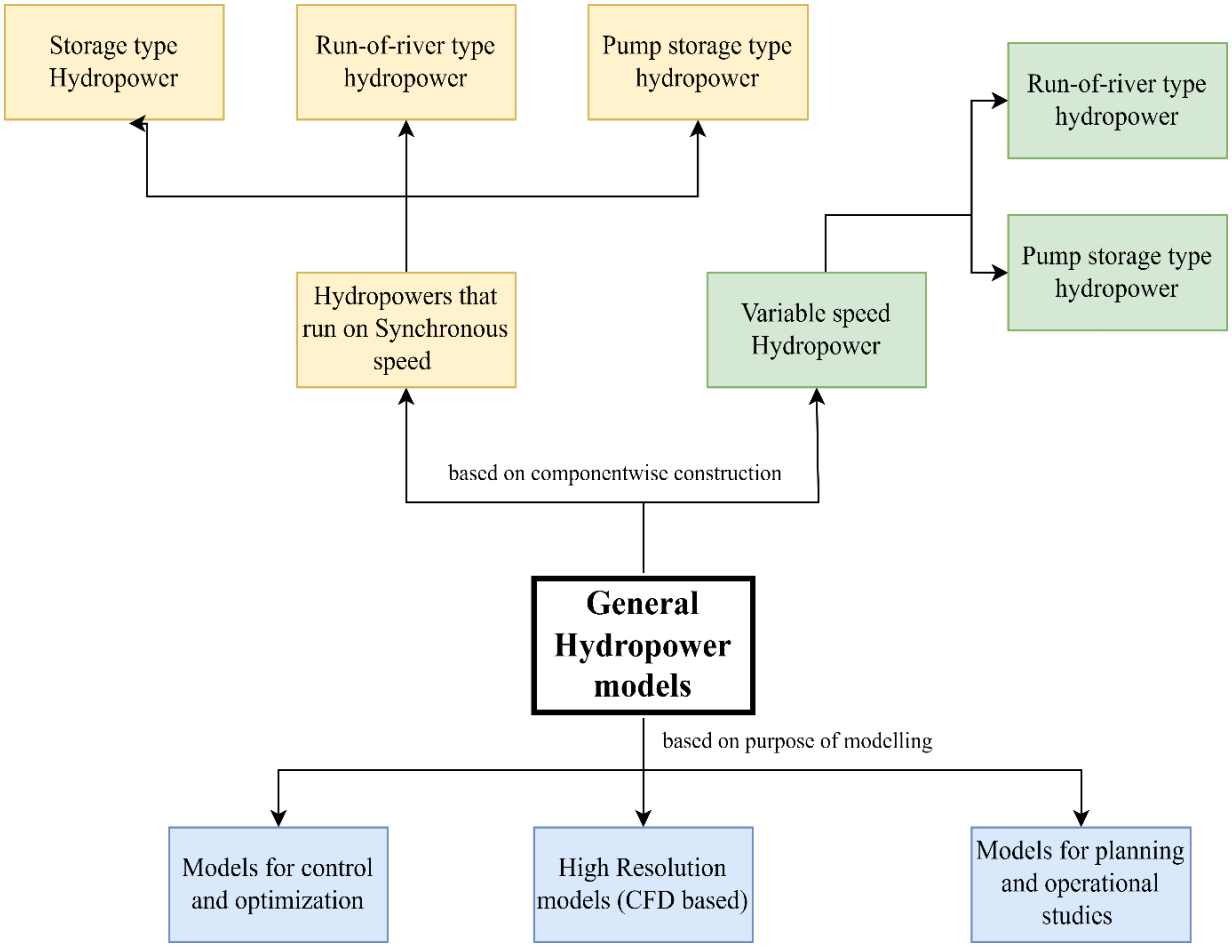


Figure 1 Classification of hydropower models

a whole hydropower system, but most of the modelling has been done focusing the kind of studies involved, making the task of a particular field easier.

A simple classification of the hydropower models is shown in Figure 1. The upper half of the figure shows the classification in terms of the components used. The hydropower plants that run with constant speed and variable speed are differentiated based on the availability of converters or governors for simple understanding. Also, both of these kinds can further be looked as Run of River (RoR), storage or pump storage type. Storage type of hydropower plants are less likely to be operated as variable speed and so, not shown in the classification of (Variable Speed Hydropower) VSHPs(Nobile, Sari and Schwery, 2018b; Valavi and Nysveen, 2018a; T. I. Reigstad and Uhlen, 2020)The lower half of the same figure classifies the available hydropower models based on their purpose of modelling. The models may have different requirements to enable us perform different analysis. For example, models for control and optimization ought to represent the dynamics of the system as accurately as possible, the models associated with planning are less concerned with dynamics and are more concerned with the amount of power produced, its availability, environmental constraints, etc. Computational Fluid Dynamics (CFD) models do not usually deal with other dynamics of the hydropower system as a whole but look at the fluid dynamics and its details for different components like turbines, guide vanes, etc. Models made for condition monitoring are concerned with the sensor data and its analysis (Z. Wenjing, 2017; Liu *et al.*, 2019; Tor Inge Reigstad and Uhlen, 2020a; Kerdphol *et al.*, 2021; Reigstad and Uhlen, 2021; Kincic *et al.*, 2022; Sapkota *et al.*, 2022)

This paper attempts to present a brief review of the hydropower models found widely in the literature by classifying them into three major categories which are (i) Models for Control and Optimization, (ii) High Resolution models (CFD models) and (iii) Models for planning and Operational studies. Although we may derive specific sub-categories even under these three major categories and critically review each of them, this paper is aimed at studying the ‘Models for control and Optimization’ in detail and will present only a brief overview and major modelling gap for the other two.

2 General Overview

In representing a hydropower system, the physical phenomenon behind the working of each component of a hydropower is necessary to understand. Firstly, the major components present in a hydropower which show the approximate path for the flow of power and energy conversion units like turbine and generator is discussed. Figure 3 shows a block diagram representation of a hydropower plant, trying to highlight differences in the major components present in a VSHP and a normal power plant. The governing principles for each component is briefly discussed in this section.

- Upstream flow region (conduits and reservoirs): The first component in a hydropower model is upstream water flow region which connects the water flow from reservoir outlet to the surge tank (a kind of reservoir). This consists of pipes and reservoirs which are modelled based on different conservation principles in physics. Equations for mass and momentum conservation describe the flows in conduits, dynamic equilibrium requires the satisfaction of Newton's second law of motion and the condition of continuity requires that the available space inside a conduit be occupied by water at all times (de Mello *et al.*, 1992; Alexandra and Tenorio, 2010).
- Penstock: Penstock involves basic phenomenon of momentum balance and the mass is considered constant in a closed conduit like penstock. Also, penstock modelling is associated with water pressure balance due to gate closing and opening action(de Mello *et al.*, 1992; K. Nabd and A. Jesus Fraile, 2017).
- Turbine: Hydraulic turbines derive mechanical power from the force exerted by water as it falls from an altitude. May it be impulse or reaction turbines, the mechanical power developed by the turbine is usually dependent on the flow rate, head and the efficiency and is modelled based on law of conservation of energy in general(P. Kundur, 2009).
- Generator: Generator used in hydropower plants has two parts namely: stator (the stationary part) and rotor (the rotating part). The stator is generally represented using Park transformation which gives the equations for d-axis and q-axis parameters. And the rotor is represented using a second order differential equation known as Swing equation which relates the Power output from the generator with the rotation(P. Kundur, 2009).
- Converter: These are used in Variable speed Power plants and are modelled using power electronic components like Thyristor or IGBT with controlled switching. The control unit in switching action is designed to maintain the power output from VSHPs to have same frequency as the grid (Tiwari, Nilsen and Mo, 2021). These have been undergoing refinement for better performance lately.
- Governor: Governors for hydropower plants work on two basic principles, namely mechanical hydraulic action or electronic action. Mechanical hydraulic governors work by displacing the fluid and moving the piston and electronic governors generally work on PID control action. The governing mechanism of hydropower is evolving towards robust control strategies using adaptive and predictive control algorithms(Li and Zhou, 2011; Guo and Yang, 2018).

3 Research review

3.1 Models for Control and Optimization

Modelling the power plant for dynamic studies have been carried since many decades and has still been undergoing improvement. The following two sub-groups categorize the hydropower models available based on their speed.

3.1.1 Hydropower's that run on synchronous speed

A work by 'IEEE working group on Prime mover and energy supply models for system dynamic performance studies' in 1992 marks the framework for hydraulic turbine and its control models are practised until now (de Mello *et al.*, 1992). Authors in this work have developed a non-linear turbine model assuming a non-elastic water column in penstock which is linearized about an operating point later on. The effect of friction losses is also included in this representation. The linearized models are claimed to be useful in the studies of control system using linear analysis tools like frequency response, eigenvalue, etc. however, the non-linear models are required for large disturbance studies and large transients.

This is the baseline for the models working with prime movers including water supply conduit and prime mover speed controls. Figure 2 shows the general relationship among mechanically rotating parts and the water flow channels of a hydropower plant which controls the dynamics of a hydropower plant. Many recent graduate and doctoral thesis works have used this representation and baseline and proposed further improvements in the models (Alexandra and Tenorio, 2010; Splavska, 2017a; Z. Wenjing, 2017; Reigstad Tor Inge, 2021).

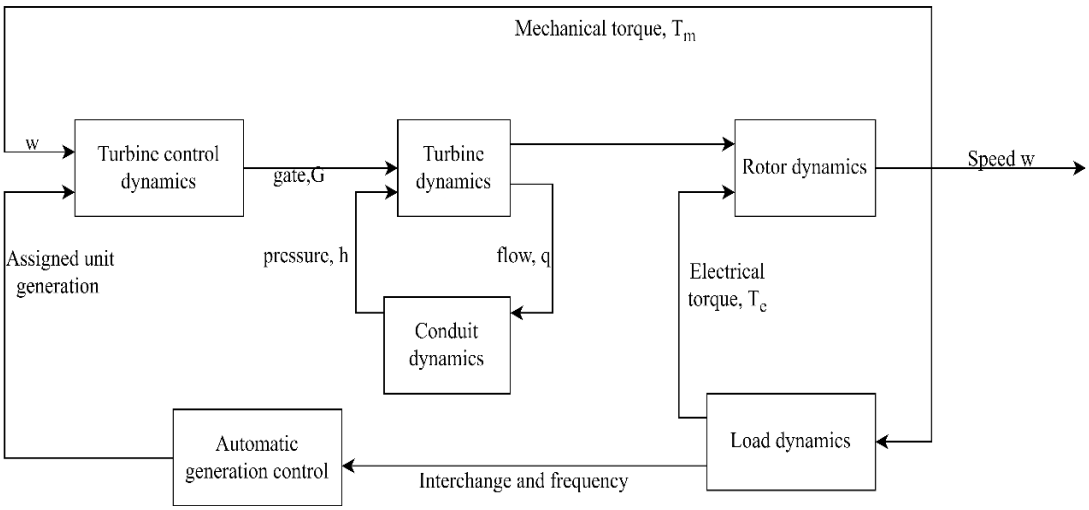


Figure 2 Block diagram representation of hydro-prime mover system and controls(de Mello *et al.*, 1992)

The earlier models were considered adequate for typical first swing stability simulations, but more issues like longer transient stability problems simulation, low frequency oscillations, islanded operation, load rejection, system restoration, water hammer dynamics, pump storage generation with complex hydraulic structures, etc. had to be addressed with the wide practise of using hydropower (de Mello *et al.*, 1992; Fang *et al.*, 2008; Acakpovi, Hagan and Fifatin, 2014; Yang *et al.*, 2015; Guo and Yang, 2018; Rheinheimer *et al.*, 2023).

For the case with long penstocks, the pressure differences and water compressibility generate significant dynamic behaviour which must be taken into modelling consideration. The water pressure is assumed to be analogous to sound waves propagation in water and the wave propagation principle is used to model the long penstocks in (de Mello *et al.*, 1992) which is termed as method of characteristic modelling in (Alexandra and Tenorio, 2010). This wave propagation model introduces a tan hyperbolic function to represent the water hammer effect in the long penstocks. This makes the functioning of the penstock non-linear. Also, the authors in (Alexandra and Tenorio, 2010) use two more methods to model the penstock, namely: Finite volume method (FVM) and electrical circuit equivalent method. FVM is associated with discretizing the main governing equations making it representable in suitable PDEs form to apply Model predictive control (MPC) algorithms. Electrical circuit equivalent method says that the flow of water is analogous to flow of current in a RLC circuit. The pressure flow is assumed analogous to travelling waves in transmission lines and the equivalent R, L and C values are derived based on penstock parameters. Most of the research until now use these methods to model the water hammering in penstocks while studying the dynamic behaviour of hydropower (P. Kundur, 2009; H. Ardul Munoz, M. Petrous and J. Dewi Ieuan, 2013; Li *et al.*, 2016; K. Nabd and A. Jesus Fraile, 2017; Guo and Yang, 2018; Cassano *et al.*, 2021; Reigstad Tor Inge, 2021; Zhang *et al.*, 2022). Also, multiple penstocks supplied from a common water column are modelled just based on the flow in the upper manifold to be equal to the flow in each penstock and the governing principles remain the same (de Mello *et al.*, 1992; H. Ardul Munoz, M. Petrous and J. Dewi Ieuan, 2013). Based on water hammer consideration in penstock modelling as described in (de Mello *et al.*, 1992; P. Kundur, 2009; H. Ardul Munoz, M. Petrous and J. Dewi Ieuan, 2013; K. Nabd and A. Jesus Fraile, 2017; Cassano *et al.*, 2021; Zhang *et al.*, 2022), the hydropower models are classified as the ones with ‘elastic water column models’ and ‘inelastic water column models’ which have significant differences in dynamic the response of hydropower plants.

One more significant component in the upstream flow region is the surge tank which is proven to be of utmost importance while studying the dynamic behaviour of hydropower plants. This is why the presence or absence of surge tank brings huge difference in planning the control of hydropower plant (P. Kundur, 2009; H. Ardul Munoz, M. Petrous and J. Dewi Ieuan, 2013). The surge tank model is derived from the continuity of flow at the two junctions which can further consider pressure balance, mass balance, momentum balance and forces acting on the surge tank(Alexandra and Tenorio, 2010; Pandey and Lie, 2021; Reigstad Tor Inge, 2021; Pandey *et al.*, 2022). Time domain models and s-domain models of surge tanks are in wide practice for the models used for control. (de Mello *et al.*, 1992; Fang *et al.*, 2008; Alexandra and Tenorio, 2010; H. Ardul Munoz, M. Petrous and J. Dewi Ieuan, 2013) present the s-domain model of surge tank which approximates the storage constant of surge tank and predict the dampening of water hammer in the penstock while the authors in(Splavska, 2017a; Pandey and Lie, 2021; Pandey *et al.*, 2022) present time domain model of the

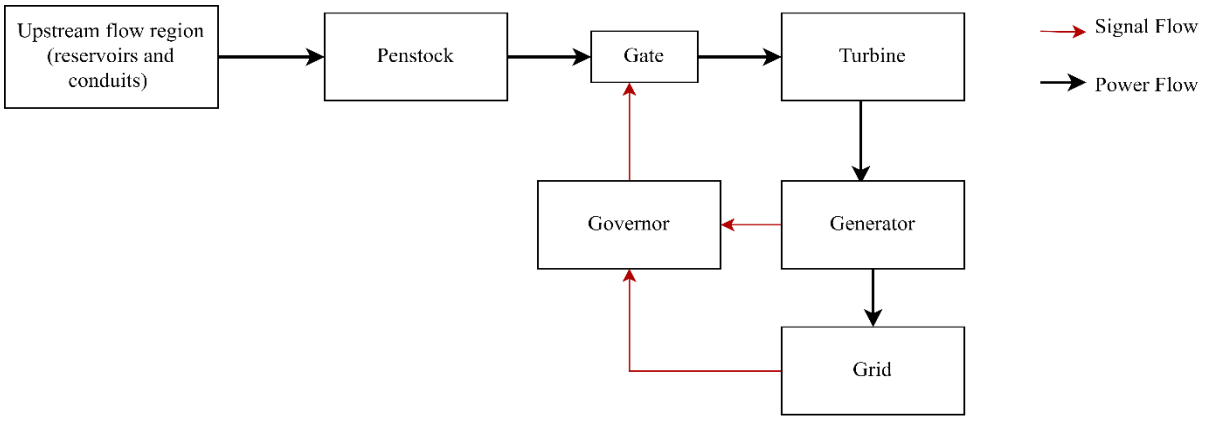


figure a

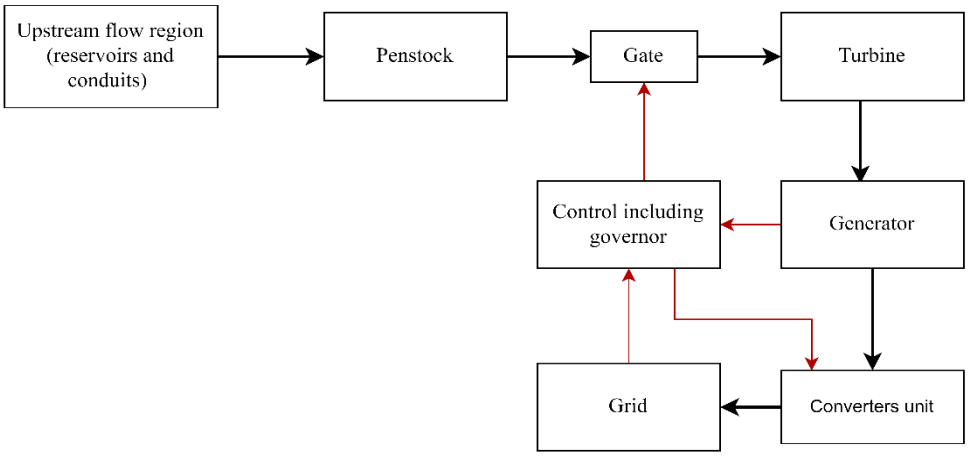


figure b

Figure 3 Block diagram representation of (a) General Hydropower (b) Variable Speed Hydropower

surge tank which is represented as a differential equation. (Pandey and Lie, 2021; Pandey *et al.*, 2022) have developed the models' representation in modellica and is available as open HPL.

Turbine modelling is necessary because the hydraulic turbine dynamics seem to have considerable effect on power system stability. As can be seen form Figure 2 the turbine dynamics have direct influence on rotor dynamics of generator. Authors in(P. Kundur, 2009) talk about grid stability in terms of rotor angle, which directly varies according to the turbine rotation dynamics. The functioning of turbine is non-linear because of the action of water pressure and head on it. The major governing principle behind turbine modelling is the balance of mechanical power represented by the equation

$$P = \eta * \rho * q * g * h$$

where P=Mechanical Power generated, ρ=Water density, Q= discharge, g = acceleration due to gravity and h = head.

Hydro turbines have non-linear performance in reality but are studied by linearizing the mathematical models as well. (Cassano *et al.*, 2021) presents the performance of linearised models of hydropower which implements linearised Francis and Kaplan turbine equations. Linearization based on first order Taylor expansion is claimed to give more tractable alternative to non-linear models and also can be used for model predictive control based on convex optimization.

Authors in (P. Kundur, 2009) derive the transfer function of turbine with certain assumptions which falls under the linearized model type. The non-linear behaviour is carried by the assumption that mechanical power is directly proportional to the square root of the head. Also, authors in the same book define water starting time as the amount of time required for a head to accelerate water in the penstock from standstill to the rated value, which is one of the important factors in turbine dynamic-study. This helps us predict the change in water flow and pressure with respect to unit gate opening. The usage of water starting time is widely found in turbine modelling studies as in (Kishor, Singh and Raghuvanshi, 2007; Alexandra and Tenorio, 2010; Zhou, Lie and Glemmstad, 2011; Yang *et al.*, 2015;

Li *et al.*, 2016; K. Nabd and A. Jesus Fraile, 2017; Guo and Yang, 2018; Reigstad Tor Inge, 2021). Furthermore, (de Mello *et al.*, 1992) is one of the earliest literature making the baseline for modern turbine dynamic models which starts with non-linear modelling and ends with the linearized turbine models about an operating point assuming both elastic and inelastic water columns. This concept has been studied until today (Kishor, Saini and Singh, 2007; Acakpovi, Hagan and Fifatin, 2014; K. Nabd and A. Jesus Fraile, 2017; Cassano *et al.*, 2021; Gao *et al.*, 2021). The non-linear models of turbine are further represented in the form of ordinary or partial differential equations which make it easier to implement the modern control methods. Linearized turbine models are seen as a set of six partial differential equations(Z. Wenjing, 2017) while non-linear turbines are modelled using Euler's equations. (Nielsen, 2015; Splavska, 2017b; Reigstad Tor Inge, 2021). Also, some authors model turbine as simple non-linear function in the form of analytic expression as a function of wicket gate, water head and speed (Li and Zhou, 2011; Li *et al.*, 2016).

The generators used in hydropower are synchronous generators. The modelling of synchronous machines has been worked on and undergone considerate degree of improvement back in 1920s and 1930s itself (P. Kundur, 2009). It is modelled using Park's transformation to represent voltage equations in the form of d-axis and q-axis parameters. Then comes the rotor swing equation which relates electrical power and rotor swing (P. Kundur, 2009; H. Ardul Munoz, M. Petrous and J. Dewi Ieuan, 2013; Guo and Yang, 2018; Brezovec, Kuzle and Krpan, 2022). Generator models are more widely used in transient stability studies for power system to test the system's robustness against electrical and mechanical faults(P. W. Sauer and M. A. Pai, 2006; HM Gibson *et al.*, 2019; Brezovec, Kuzle and Krpan, 2022). Furthermore, standstill frequency response (SFR) and rotating dynamic response (RTDR) are two methods currently used to derive the generator parameters. These provide further flexibility and enables to handle the non-linearities with less computational burden(Fonseka, de Silva and Dong, 2021). Also, some generators used in VSHPs are permanent magnet synchronous generators (PMSG) which requires the representation of stator equations in dq form (Gao *et al.*, 2021).

Despite of wide availability of hydropower models for control and application of numerous control algorithms, there is always a room for improvement. The models implementing robust control, most preferably the ones with time-domain dynamic representation have been worked upon very less. Though some time-domain models for synergetic control studies and state space models of hydropower and (Pump Storage Hydropower) PSH have been explored in (Zhang *et al.*, 2022) and (Dong *et al.*, 2020), transient studies for grid side have been studied more than for the generator side. With the concern for reliable and resilient power system, the control for hydropower must be adequate from all the possible aspects. This demands for more accurate dynamic models and even undertaking probabilistic dynamics as well.

3.1.2 Variable Speed Hydropower (VSHP)

Development of variable speed plants dates back to early 1990s in Japan (Valavi and Nysveen, 2018b). Since then, VSHPs have been undergoing pioneering achievements and becoming popular because of their capability to provide additional ancillary services to the grid apart from power production. Figure 3 represents the major components reflecting the difference between normal hydropower plants and variable speed hydropower plants. VSHP consist of a converter in addition to a normal hydropower component which are either operated as full size or partially when needed. The converters' main task is to fed the power to the grid maintaining constant frequency despite of changing generator speed (Tiwari and Nilsen, 2020; Tiwari, Nilsen and Nysveen, 2020, no date). The turbine is allowed to deviate from its normal rotating speed enabling itself to vary the output power very quickly because of the fast-acting converter technology (Nobile, Sari and Schwery, 2018b). Though popular in pump storage plants because of the ability to control frequency in pumping mode, variable speed plants are not limited to PSH only. They can also be used in HVDC connected hydropower facilities because the frequency of the generator is not tied to the grid and hence the operation of plant can be optimized by adjusting rotational speed (Camacho, 1997). Furthermore, small hydropower with considerable head and flow variations can benefit implementing variable speed operation as maintenance of high efficiency is possible (Borkowski and Majdak, 2020).

The turbine is modelled using Euler's turbine equations which have considered rotational speed along with guide vane opening to find the torque and flow in (Nielsen, 2015). A one-dimensional numerical model of Francis turbine tuned with test data for VSHP operation is presented in (Nag and Lee, 2018). With the consideration of turbine side only and utilizing the water column and reservoirs models previously presented, author in (T. I. Reigstad and Uhlen, 2020) has compared four different hydraulic models namely: Hygov model, IEEE model, Euler's model and Linearised model for VSHP operation. The paper discusses that when the models are linearised, all four models have a similar performance for governor control however Euler and IEEE models add dynamics to the penstock. This is also stated in other literatures (P.

W. Sauer and M. A. Pai, 2006; P. Kundur, 2009; Nielsen, 2015). A simple turbine model does not consider the relationship of turbine efficiency with rotational speed, which is crucially important in VSHP turbines. So, Euler's model is considered as most suitable for simulating transients and variation in rotational speed in VSHP which have significant impact in fast frequency response of gird (Nielsen, 2015; Nobile, Sari and Schwery, 2018b; Tor Inge Reigstad and Uhlen, 2020a; Reigstad and Uhlen, 2021). But the interaction of this model with the power system also needs to be studied before considering it to be the most suitable one.

The other major component of VSHP is the converter. Authors in (Tiwari, Nilsen and Nysveen, 2020) and (Tiwari, Nilsen and Nysveen, no date) talk about the converters for variable speed pump storage power plants. Neutral point converters (NPC), multilevel converters (MMC) and full sized active neutral point converters (ANPC) have been in application for both synchronous machines and doubly fed induction machine (DFIC) in pump storage plants (Tiwari, Nilsen and Nysveen, no date; Tiwari, Nilsen and Mo, 2021). ANPC are claimed to have high starting torque which is essential requirement for machine side application in VSHP application but is threatened by the converter losses (Tiwari and Nilsen, 2020). Precise control strategies like hierarchical control and optimization algorithms, model predictive control, stator flux regulatory control, etc. for these converters have been discussed in many literatures such as in (J. Kristansen Noland, J. Hagset and Stavnesli, no date; S. K. Peter *et al.*, 2014; Tor Inge Reigstad and Uhlen, 2020b; Tiwari, Nilsen and Mo, 2021), but the models for coordinated control are still lacking.

Research have been performed to improvise the models for more accurate representation of power plant dynamics both from load side and the turbine side since many decades. Also, the grid side interactions like frequency reserves, ROCOF, transient analysis, etc. have been studied a little, but the uncertainties that might occur during the plant operation, particularly for VSHPs have still been left behind. With the prevailing examples of grid failures because of changing environmental conditions and other known or unknown uncertainties, there is a dire need to develop models which can represent these environmental and other disturbances threatening the grid. The probabilistic disturbances can be modelled and added to the existing models. This enables to come up with the control plans in cases of such unprecedented conditions. For robust control of hydropower plants, the models used for designing these controllers should be able to reflect the effect of uncertainties on the system. Thus, describing the uncertainties during the modelling phase is needed. Furthermore in (Dong *et al.*, 2020) a concept called quaternary PSH which involves bifurcated penstock system is introduced, but description of a coordinated control of multiple hydropower plants operating from a single water channel is not yet available in open literature.

3.2 High Resolution Models (CFD)

Computational Fluid Dynamics (CFD) is used to look at the effect of fluid dynamics on several component of the power plant. The most extensively studied component in CFD analysis is the turbine (Tiwari *et al.*, 2020). Authors in (Lain and Mejia, 2022) claim that physical modelling have gradually been replaced by CFD modelling techniques which are used to study hydro-kinetic along with hydro-dynamic studies. With the increasing popularity of PSH as energy storage elements and a good ancillary service provider for electric grid, CFD simulation of model turbine as a pump has been performed in (Deng *et al.*, 2022) to quantify energy loss and entropy generation. Furthermore, studies about change in guide vane air foil on the flow characteristics of draft tube for improvement of energy recovery, vibrations intensity, stable operation of turbine, design methods of multiphase pumps for hydrodynamic and structural points of view, etc. have been performed widely in the past (Benavides-Morán, Rodríguez-Jaime and Laín, 2022; Niebuhr *et al.*, 2022; Peng *et al.*, 2022).They conclude that the power coefficient is affected by presence of free surface. Similarly, authors in (Lopez Mejia *et al.*, 2021) propose practise guidelines for CFD simulations in turbines studying the performance standard of vertical axis and horizontal axis turbines and those in (Xiong, Deng and Chen, 2021) study about flapping motion in tail edge which is found to present a better stability for turbine. Although it is said that high level high-order models like CFD are not practical in modelling the whole hydropower, a few low order models are studied for the purposes like fault occurrence, estimation of number and geometry of components for hydropower, etc in practice(Li *et al.*, 2021; YoosefDoost and Lubitz, 2021; Saeed *et al.*, 2022). Furthermore, optimization of hydraulics for a Kaplan turbine at different operating conditions is studied in (Benigni *et al.*, 2014) and also the curved paddles in the wheel enhanced generation by 10-20 % is described in (Akinyemi and Liu, 2015).

CFD models mainly deal with component-wise performance in detail rather than looking at the whole hydropower. With the evidence that CFD models are mainly concerned with turbine study, we still find a gap in the study of small details like labyrinth seals study, pressure balance in draft tube for Kaplan turbines, etc.

3.3 Models for Planning and Operational Studies

The operation modes of hydropower are widely varying in recent times and so the operating models are pushed toward the boundary. Hydropower models developed for operation and planning studies generally consider nominal water availability only and the environmental constraints like headwater limits variation, average flow variation, etc are not incorporated in the models. This might lead to inaccurate long term and short-term system studies like erroneous transmission flow and response to contingencies. A few gaps in the order of priority have been identified which are listed as follows(Kincic *et al.*, 2022)

- Water availability not properly represented in system models
- Interdependencies among hydro projects and environmental constraints are not properly represented in system models
- Rough zones are not represented in the power system model so generation dispatch in system studies might not be realistic
- Many dynamic models of hydro generation turbines are outdated
- Inaccuracy in frequency response during simulation studies
- Data issues and incorrect parameters values in dynamic models
- Advanced pumped storage models are not widely available

Each of these gaps have been critically analysed and several ways to model the hydropower have been suggested more accurate results. Water availability issue stated above can be addressed by collecting more precise and granular data, representation of constraints within production cost models and capacity expansion models with watershed models. The authors in (Kincic *et al.*, 2022) further state that the base case models also known as power flow models used for power system studies only maximum generated power. However, this might not always be the case as the generation keeps varying with seasonal water variations. Also, authors in the same report come up with the fact the water levels affect the droop and governor response as well which might lead to over representation of turbine-governor response to system frequency events, voltage stability, and transient stability issues. The simulation studies have been done in the HYGOV4 governor dynamic model implemented with a gas turbine which shows that 5-10% variation in head value can significantly affect the dynamic response and frequency recovery of the source.

Water basins have different interdependencies like tailwater and forebay level change rate limits for flood regulation, effect on aquatic ecosystem, water ratio maintenance for cascaded plants which are required to be coupled in software but are not.

Authors in (Dong *et al.*, 2019) talk about problems in reliability because of oscillation phenomena in the hydropower operating in rough zones. The turbine undergoes a mechanism called vortexing which leads to oscillations in power systems, but there is no knowledge of these restricted zones (that lead to oscillations) of operation in power system simulation studies.

Incorrect parameters values in the models also have a huge role in generating inaccurate results for the planning studies. This can lead to the models being too optimistic or too pessimistic. Talking about dynamic performance as stated in (Pereira *et al.*, 2003; Kou *et al.*, 2016; Soni, 2016), the incorrect parameters used for governor-generator modelling has impacts on turbine gain constant and frequency of the system. Turbine gain constant is directly associated with the mechanical power generated by the turbine which is one of the key parameters for grid studies ranging from stability to planning. Furthermore, the same issue with data leads to false frequency nadirs which projects wrong frequency response in the grid (Pereira *et al.*, 2003; Soni, 2016). This is a huge threat to the grid operating stability and for decision making regarding expansion planning, contingency and line flows.

Pump storage hydropower (PSH) plant is already a mature energy storage technology but there still exist gaps in developing advanced PSH models which anticipate the real-time operation on PSH. In many existing software, the pumping mode of a PSH is modelled as motor and generating mode is developed as a hydro-generator but the transition between them is often ignored (Kincic *et al.*, 2022). Models in generating and pumping modes are different and need to take into consideration the water hammer, throttling of the wicket gate for pump starting and shutting down, etc(Nobile, Sari and Schwery, 2018a). Furthermore, adjustable speed PSH can out space conventional PSH which may lead to huge market growth and installation (Valavi and Nysveen, 2018a). User defined models have been developed in (J. Feltes *et al.*, 2013) which try to resolve the above stated issues but these models have not been validated and commercialized. As seen from the models with operation and planning studies, the deterministic approaches for grid resiliency and reliability have been performed both from grid and load side. Moreover, much attention has been given to the load models and market operation strategies. But there is a need to conduct studies from probabilistic approach as well. The unanticipated changes in load or operating conditions of hydropower have not been taken into consideration much.

A simplified brief about the development of hydropower models from the past to the present is shown in Table 1.

Table 1 Hydropower models: chronological brief

Classifications	1992-2010 (Past)	2010-2023(current)	Future Prospects
Models for Control and Optimization	- Basis for hydro-turbine control models for dynamic studies applied until present context - Dynamic models mathematics for transient and control study - Grid support using hydropower - Studies in frequency domain, mainly using classical control methods	-Concept of variable speed hydropower emerged - Intermittency of renewables balancing using hydropower -Importance of grid support using hydropower and pump storage flourished -Dynamic models for transient and control study with increased detail in models -Studies both in s-domain and t-domain along with the application of robust control	- Dynamic models which can represent uncertainties - Models representing multiple hydropower in same channel for control studies
High resolution models (CFD)	- Turbine models for cavitation studies - Physical models gradually replaced by CFD models	- Turbine models covered the area of pump turbines as well - variable speed model components study - Condition monitoring of the components	- Labyrinths seals' study - more insights on variable speed power plants
Models for planning and operational studies	- Nominal steady state operations considered - Less attention on environmental constraints - Electricity market operation strategies	-Consideration of load models -Use of AI for numerical models -Reliability studies -Electricity market deregulation and flexibility	- More focus on probabilistic reliability studies along with the ongoing deterministic

4 Conclusion

This paper presents a surficial picture about the existing hydropower models, linking the commonly seen classification to the classification of models for purpose. The types of models based on purpose of modelling have been stated as: Models for control, models based on CFD and models for operation and planning.

The paper provides more broader overview about models for control which are found to be developed on the basis of principles of physics like mass, continuity and energy balance representing the dynamics of the system associated. Despite of having the same mathematics, way of presenting the models varies as per the requirement. For example: the non-linearity brought about by the water pressure behaviour in the penstock is ignored in some models whereas considered to be important in some other models. Also, the models developed for modern control implementation are found to be more detailed and those involving non-linear optimization are preferred to be developed in time-domain. The dynamic models developed are also popularly used for transient studies. Furthermore, VSHP models also have been developed for control purpose among which PSH models are more popular. The converters implemented in VSHPs have become very popular for their ability to change the output power very fast with the changing speed capability which can improve the frequency response in grid.

Moreover, considerable amount of work has been done in CFD modelling and the models for planning studies. The model developed using CFD re also used for control, but most of its application has been found in the component wise analysis and a single component control rather than coordinated control. Also, models for operating and planning are mostly used to study about the grid impact and environmental impact. The scope extends a bit to economy and the society as well.

With the study of the available models and their application, the major work that can be done immediately is the inclusion of probabilistic analysis of uncertain events in all three fields. Focusing on control, the optimization of

performance or operation in the presence of uncertainty can be an interesting field for future research since control system can make the hydropower more capable to support the grid thereby improving the resiliency and reliability of the overall power system.

5 References

Acakpovi, A., Hagan, E. Ben and Fifatin, F.X. (2014) 'Review of Hydropower Plant Models', *International Journal of Computer Applications*, 108(18), pp. 975–8887.

Akinyemi, O.S. and Liu, Y. (2015) 'CFD modeling and simulation of a hydropower system in generating clean electricity from water flow', *International Journal of Energy and Environmental Engineering*, 6(4), pp. 357–366. Available at: https://doi.org/10.1007/s40095-015-0180-2.

Alexandra, L. and Tenorio, L. (2010) *Master of Science in Electric Power Engineering Hydro Turbine and Governor Modelling Electric-Hydraulic Interaction, Kjetil Uhlen.*

Benavides-Morán, A., Rodríguez-Jaime, L. and Laín, S. (2022) 'Numerical Investigation of the Performance, Hydrodynamics, and Free-Surface Effects in Unsteady Flow of a Horizontal Axis Hydrokinetic Turbine', *Processes*, 10(1). Available at: https://doi.org/10.3390/pr10010069.

Benigni, H. *et al.* (2014) 'Refurbishment of semi-regulated Kaplan turbines and customised hydraulic development by means of CFD', in.

Borkowski, D. and Majdak, M. (2020) 'Small hydropower plants with variable speed operation—an optimal operation curve determination', *Energies*, 13(23). Available at: https://doi.org/10.3390/en13236230.

Brezovec, M., Kuzle, I. and Krpan, M. (2022) 'Detailed mathematical and simulation model of a synchronous generator', *Journal of Energy - Energija*, 64(1–4), pp. 102–129. Available at: https://doi.org/10.37798/2015641-4147.

Camacho, J.R. (1997) 'Guide For Preliminary Design And Specification Of Hydro Stations With Hvdc Unit-Connected Generators. Analysis in the Parameterization of the Inductances and Resistance of the Rotor of a Three Phase Induction Motor using Finite Elements Method View project Parameter determination of photovoltaic sources of energy. View project'. Available at: https://doi.org/10.13140/RG.2.1.4015.1204.

Cassano, S. *et al.* (2021) 'Performance Assessment of Linear Models of Hydropower Plants', in *Proceedings of 2021 IEEE PES Innovative Smart Grid Technologies Europe: Smart Grids: Toward a Carbon-Free Future, ISGT Europe 2021*. Institute of Electrical and Electronics Engineers Inc. Available at: https://doi.org/10.1109/ISGTEurope52324.2021.9639912.

Deng, Y. *et al.* (2022) 'Research on Energy Loss Characteristics of Pump-Turbine during Abnormal Shutdown', *Processes*, 10(8). Available at: https://doi.org/10.3390/pr10081628.

Dong, Z. *et al.* (2019) 'Modelling and simulation of ternary pumped storage hydropower for power system studies', *IET Generation, Transmission and Distribution*, 13(19), pp. 4382–4390. Available at: https://doi.org/10.1049/iet-gtd.2018.5749.

Dong, Z. *et al.* (2020) 'Developing of Quaternary Pumped Storage Hydropower for Dynamic Studies', *IEEE Transactions on Sustainable Energy*, 11(4), pp. 2870–2878. Available at: https://doi.org/10.1109/TSTE.2020.2980585.

Fang, H. *et al.* (2008) 'Basic modeling and simulation tool for analysis of hydraulic transients in hydroelectric power plants', *IEEE Transactions on Energy Conversion*, 23(3), pp. 834–841. Available at: https://doi.org/10.1109/TEC.2008.921560.

Fonseka, P.A.J., de Silva, S. and Dong, Z.Y. (2021) 'A Framework to Derive Synchronous Generator Modeling Parameters from Grid Connection Tests', *Electric Power Components and Systems*, 49(20), pp. 1525–1535. Available at: https://doi.org/10.1080/15325008.2022.2134511.

Gao, J. *et al.* (2021) 'Variable-Speed Hydropower Generation: System Modeling, Optimal Control, and Experimental Validation', *IEEE Transactions on Industrial Electronics*, 68(11), pp. 10902–10912. Available at: https://doi.org/10.1109/TIE.2020.3031528.

Guo, W. and Yang, J. (2018) 'Modeling and dynamic response control for primary frequency regulation of hydro-turbine governing system with surge tank', *Renewable Energy*, 121, pp. 173–187. Available at: https://doi.org/10.1016/j.renene.2018.01.022.

H. Ardul Munoz, M. Petrous and J. Dewi Ieuan (2013) *Modelling and Controlling hydropower Plants.* springer.

HM Gibson *et al.* (2019) 'Modelling of Synchronous Generator for Transient Stability in Power System', in *2019 2nd International Conference on High Voltage Engineering and Power Systems (ICHVEPS)*. IEEE, pp. 125–129. Available at: https://doi.org/10.1109/ICHVEPS47643.2019.9011059.

J. Feltes *et al.* (2013) *Modeling Adjustable Speed Pumped Storage Hydro Units Employing Doubly-Fed Induction Machines Decision and Information Sciences*. Argonne, USA. Available at: www.anl.gov.

J. Kristansen Noland, J. Hagset and Stavnesli (no date) 'Stator Flux-Regulatory Excitation Control inConverter-Fed Synchronous Machines forPumped-Storage Variable-Speed Hydropower'.

K. Nabd and A. Jesus Fraile (2017) *IET ENERGY ENGINEERING SERIES 100 Modeling and Dynamic Behaviour of Hydropower Plants*.

Kerdphol, T. *et al.* (2021) 'An Overview of Virtual Inertia and Its Control', in *Power Systems*. Springer Science and Business Media Deutschland GmbH, pp. 1–11. Available at: https://doi.org/10.1007/978-3-030-57961-6_1.

Kincic, S. *et al.* (2022) *Hydropower Modeling Gaps in Planning and Operational Studies*. Available at: http://www.ntis.gov.

Kishor, N., Saini, R.P. and Singh, S.P. (2007) 'A review on hydropower plant models and control', *Renewable and Sustainable Energy Reviews*, pp. 776–796. Available at: https://doi.org/10.1016/j.rser.2005.06.003.

Kishor, N., Singh, S.P. and Raghuvanshi, A.S. (2007) 'Adaptive intelligent hydro turbine speed identification with water and random load disturbances', *Engineering Applications of Artificial Intelligence*, 20(6), pp. 795–808. Available at: https://doi.org/10.1016/j.engappai.2006.11.014.

Kou, G. *et al.* (2016) 'Impact of Governor Deadband on Frequency Response of the U.S. Eastern Interconnection', *IEEE Transactions on Smart Grid*, 7(3), pp. 1368–1377. Available at: https://doi.org/10.1109/TSG.2015.2435258.

Kumari Rupesh *et al.* (2019) 'Development Trends and Future Prospects of Hydro-Turbine Control Systems', in *IEEE Industry Applications Society Annual Meeting*.

Lain, S. and Mejia, O.D.L. (2022) 'Special Issue on "CFD Modelling and Simulation of Water Turbines"', *Processes*. MDPI. Available at: https://doi.org/10.3390/pr10112410.

Li, C. *et al.* (2016) 'Parameter identification of a nonlinear model of hydraulic turbine governing system with an elastic water hammer based on a modified gravitational search algorithm', *Engineering Applications of Artificial Intelligence*, 50, pp. 177–191. Available at: https://doi.org/10.1016/j.engappai.2015.12.016.

Li, C. and Zhou, J. (2011) 'Parameters identification of hydraulic turbine governing system using improved gravitational search algorithm', in *Energy Conversion and Management*. Elsevier Ltd, pp. 374–381. Available at: https://doi.org/10.1016/j.enconman.2010.07.012.

Li, X. *et al.* (2021) 'Effect of Tip Clearance Size on Tubular Turbine Leakage Characteristics', *Processes*, 9(9). Available at: https://doi.org/10.3390/pr9091481.

Liu, H. *et al.* (2019) 'A high-resolution hydro power time-series model for energy systems analysis: Validated with Chinese hydro reservoirs', *MethodsX*, 6, pp. 1370–1378. Available at: https://doi.org/10.1016/j.mex.2019.05.024.

Lopez Mejia, O.D. *et al.* (2021) 'Comparison of Sliding and Overset Mesh Techniques in the Simulation of a Vertical Axis Turbine for Hydrokinetic Applications', *Processes*, 9(11). Available at: https://doi.org/10.3390/pr9111933.

de Mello, F.P. *et al.* (1992) 'Hydraulic turbine and turbine control models for system dynamic studies working group on prime mover and energy supply models for system dynamic performance studies', *IEEE Transactions on Power Systems*, 7(1), pp. 167–179. Available at: https://doi.org/10.1109/59.141700.

Nag, S. and Lee, K.Y. (2018) 'DFIM-Based Variable Speed Operation of Pump-Turbines for Efficiency Improvement', *IFAC-PapersOnLine*, 51(28), pp. 708–713. Available at: https://doi.org/https://doi.org/10.1016/j.ifacol.2018.11.788.

Niebuhr, C.M. *et al.* (2022) 'Development of a Hydrokinetic Turbine Backwater Prediction Model for Inland Flow through Validated CFD Models', *Processes*, 10(7). Available at: https://doi.org/10.3390/pr10071310.

Nielsen, T.K. (2015) 'Simulation model for Francis and Reversible Pump Turbines', *International Journal of Fluid Machinery and Systems*, 8(3). Available at: https://doi.org/10.5293/IJFMS.202015.8.3.169.

Nobile, E., Sari, G. and Schwery, A. (2018a) 'Variable Speed Hydro Pumped Storage as Flexible Enabler of Intermittent Renewable Penetration', in *IEEE Power and Energy Society General Meeting*. IEEE Computer Society. Available at: https://doi.org/10.1109/PESGM.2018.8586238.

Nobile, E., Sari, G. and Schwery, A. (2018b) 'Variable Speed Hydro Pumped Storage as Flexible Enabler of Intermittent Renewable Penetration', in *IEEE Power and Energy Society General Meeting*. IEEE Computer Society. Available at: https://doi.org/10.1109/PESGM.2018.8586238.

P. Kundur (2009) *Power System Stability And Control P Kundur (2009, Mcgraw Hill), MCgrawHIll*.

P. W. Sauer and M. A. Pai (2006) *POWER SYSTEM DYNAMICS AND STABILITY*. 1st edn. Stipes Publishing Co.

Pandey, M. *et al.* (2022) 'Mechanistic Model of an Air Cushion Surge Tank for Hydro Power Plants', *Energies*, 15(8), p. 2824. Available at: https://doi.org/10.3390/en15082824.

Pandey, M. and Lie, B. (2021) 'Mechanistic modeling of different types of surge tanks and draft tubes for hydropower plants', in, pp. 131–138. Available at: https://doi.org/10.3384/ecp20176131.

Peng, C. *et al.* (2022) 'A Method for the Integrated Optimal Design of Multiphase Pump Based on the Sparse Grid Model', *Processes*, 10(7). Available at: https://doi.org/10.3390/pr10071317.

Pereira, L. *et al.* (2003) 'A new thermal governor modeling approach in the WECC', *IEEE Transactions on Power Systems*, 18(2), pp. 819–829. Available at: https://doi.org/10.1109/TPWRS.2003.811007.

Quiroga OD (2000) *Modelling and nonlinear control of voltage frequency of hydroelectric power plants.*

Reigstad, T. I. and Uhlen, K. (2020) 'Modelling of variable speed hydropower for grid integration studies', in *IFAC-PapersOnLine*. Elsevier B.V., pp. 13048–13055. Available at: https://doi.org/10.1016/j.ifacol.2020.12.2176.

Reigstad, Tor Inge and Uhlen, K. (2020a) 'Optimized Control of Variable Speed Hydropower for Provision of Fast Frequency Reserves', *Electric Power Systems Research*, 189. Available at: https://doi.org/10.1016/j.epsr.2020.106668.

Reigstad, Tor Inge and Uhlen, K. (2020b) 'Variable Speed Hydropower Conversion and Control', *IEEE Transactions on Energy Conversion*, 35(1), pp. 386–393. Available at: https://doi.org/10.1109/TEC.2019.2943233.

Reigstad, T.I. and Uhlen, K. (2021) 'Nonlinear model predictive control of variable speed hydropower for provision of fast frequency reserves', *Electric Power Systems Research*, 194. Available at: https://doi.org/10.1016/j.epsr.2021.107067.

Reigstad Tor Inge (2021) *Optimal control of variable speed hydropower Utilising model predictive control and virtual inertia for delivering power system services*. NTNU.

Rheinheimer, D.E. *et al.* (2023) 'Hydropower representation in water and energy system models: a review of divergences and call for reconciliation', *Environmental Research: Infrastructure and Sustainability*, 3(1), p. 012001. Available at: https://doi.org/10.1088/2634-4505/acb6b0.

S. K. Peter *et al.* (2014) 'Converter-Fed Synchronous machine for pumped storage hydro storage plants', in *IEEE Energy Conversion Congress and Exposition (ECCE)*. Pittsburgh, PA, USA.

Saeed, A. *et al.* (2022) 'Power Regulation and Fault Diagnostics of a Three-Pond Run-of-River Hydropower Plant', *Processes*, 10(2). Available at: https://doi.org/10.3390/pr10020392.

Sapkota, P. *et al.* (2022) 'Measurements for Condition Monitoring of Hydropower plants: A Review', in *IOP Conference Series: Earth and Environmental Science*. Institute of Physics. Available at: https://doi.org/10.1088/1755-1315/1037/1/012019.

Shahgholian, G. (2020) 'An Overview of Hydroelectric Power Plant: Operation, Modeling, and Control', *JREE*, 7(3), pp. 14–28.

Soni, S. (2016) *NERC| Reliability Guideline Power Plant Dynamic Model Verification using PMUs*.

Splavska, V. (2017a) *Simulation and control of hydro power plants*. USN.

Splavska, V. (2017b) *Simulation and control of hydro power plants*. Available at: www.usn.no.

Tiwari, G. *et al.* (2020) 'Utility of CFD in the design and performance analysis of hydraulic turbines — A review', *Energy Reports*, 6, pp. 2410–2429. Available at: https://doi.org/https://doi.org/10.1016/j.egyr.2020.09.004.

Tiwari, R. and Nilsen, R. (2020) 'Analytical Loss Equations for Three Level Active Neutral Point Clamped Converters', in *IECON Proceedings (Industrial Electronics Conference)*. IEEE Computer Society, pp. 1285–1290. Available at: https://doi.org/10.1109/IECON43393.2020.9254393.

Tiwari, R., Nilsen, R. and Mo, O. (2021) 'Control Strategies for Variable Speed Operation of Pumped Storage Plants with Full-size Converter Fed Synchronous Machines', in *IEEE Energy Conversion Congress and Exposition (ECCE)*.

Tiwari, R., Nilsen, R. and Nysveen, A. (2020) 'Active NPC Converter for Variable Speed Operation of Pumped Storage Hydropower Plant', in *IECON Proceedings (Industrial Electronics Conference)*. IEEE Computer Society, pp. 1211–1216. Available at: https://doi.org/10.1109/IECON43393.2020.9255403.

Tiwari, R., Nilsen, R. and Nysveen, A. (no date) *Evaluation and Comparison between Multilevel Converters for Variable Speed Operation of Pumped Storage Power Plants with Full-size Converters*.

Valavi, M. and Nysveen, A. (2018a) 'Variable-Speed Operation of Hydropower Plants: A Look at the Past, Present, and Future', *IEEE Industry Applications Magazine*, 24(5), pp. 18–27. Available at: https://doi.org/10.1109/MIAS.2017.2740467.

Valavi, M. and Nysveen, A. (2018b) 'Variable-Speed Operation of Hydropower Plants: A Look at the Past, Present, and Future', *IEEE Industry Applications Magazine*, 24(5), pp. 18–27. Available at: https://doi.org/10.1109/MIAS.2017.2740467.

Xiong, P., Deng, J. and Chen, X. (2021) 'Performance Improvement of Hydrofoil with Biological Characteristics: Tail Fin of a Whale', *Processes*, 9(9). Available at: https://doi.org/10.3390/pr9091656.

Yang, W. *et al.* (2015) 'A mathematical model and its application for hydro power units under different operating conditions', *Energies*, 8(9), pp. 10260–10275. Available at: https://doi.org/10.3390/en80910260.

YoosefDoost, A. and Lubitz, W.D. (2021) 'Design Guideline for Hydropower Plants Using One or Multiple Archimedes Screws', *Processes*, 9(12). Available at: https://doi.org/10.3390/pr9122128.

Z. Wenjing (2017) *Modeling, Control and Optimization of a Hydropower Plant*.

Zhang, T. *et al.* (2022) 'Synergetic Control Strategy for Water Hammer Wave and Mass Wave in Hydropower Station with Downstream Channel', *IEEE Access*, 10, pp. 81105–81122. Available at: https://doi.org/10.1109/ACCESS.2022.3195172.

Zhou, W., Lie, B. and Glemmstad, B. (2011) 'Modelling and Control of a Typical High Head Hydropower Plant in Norway', in *World Congress on Engineering and Technology* .

Traceable System of Systems Explorations Using RCE Workflows

Jorge Lovaco[a,*], Ingo Staack[b], Petter Krus[a]

[a] *Department of Management and Engineering (IEI), Linköpings university, Sweden,* [b] *Institute of Aircraft Design and Lightweight Structure (IFL), Technische Universität Braunschweig.*
jorge.lovaco@liu.se

Abstract

The System of Systems (SoS) framework plays a pivotal role in delimiting aircraft design spaces by examining interactions among its Constituent Systems (CS). Each CS has a distinct collection of capabilities, some of which may be shared with other CS. The framework explores emergent behaviours that arise from communication between the CS within the SoS. These emergent behaviours are characterized by their unattainability by any individual CS and result from their collaborative nature. The identification of these emergent behaviours enables System of Systems Engineering (SoSE) to pinpoint the most valuable configurations of the SoS, thereby maximizing the collective value. Furthermore, these emergent behaviours aid in stipulating design requirements for new systems based on the capabilities outlined in the SoS study. To map the relationship between needs, capabilities, requirements, and behaviours, maintaining traceability throughout the study is paramount.
This research employs workflows created using the Remote Component Environment (RCE), a specialized tool for structured and automated task development. The objective is to showcase RCE's integration capabilities- specifically for software tools and Python scripts- with task scheduling. This integration enables swift extraction of results, making them available at every step, thus augmenting analysis efficiency. The study focuses on the perspective of an aircraft designer during the early concept generation phase, specifically applied to the development of an electric Unmanned Aerial Vehicle (UAV) concept for wildfire detection.
Keywords: System of Systems, Aircraft Conceptual Design, Wildfire Detection, Agent-Based Simulation, Systems Engineering.

1. Introduction

The System of Systems (SoS) analysis has become an important part of systems engineering (Staack, 2019). An SoS is defined as a system that consists of several Constituent Systems (CS) and shares the following characteristic properties (Maier, 2014):

- Each of the CS can operate by itself and execute its mission as an individual.
- Each of the CS is managed independently, which means that each one of them is acquired and maintained within its own budget constraints.
- The CS are under a geographic distribution, with the distance large enough to require a non-physical communication.
- Communication and interaction between CS results in emergent behaviours, which can be not only difficult to predict but also unattainable by any single CS.
- There is an evolution throughout time of the SoS, either by adding or removing CS, upgrading or renewing them, or integrating CS different from the ones in the original composition.

Modelling and simulation become an obvious choice for finding emergent behaviours in an SoS configuration and to study its evolution when the CS arrangement changes.
With an SoS being a complex assembly of systems, large numbers of possible combinations might be found during the different analyses. Thus, the requirement spaces must be meticulously managed to prevent confusion and ensure traceability at every step of the process (Mori, 2018). Traceability ensures that the customer needs are correctly understood and linked to requirements and use cases (Luzeaux & Ruault et al., 2013). Requirement traceability also supports greatly the analysis of an SoS. It narrows down the design space and enhances trade-off analysis (Staack, 2019), from the top level of the SoS, down to the subsystem level of the CS in an asynchronous manner (Dahmann, 2008). Poor traceability makes it difficult to identify purposes or goals resulting in the less convenient "*bureaucracy-driven architectures*" (Maier, 2014). Enhancing traceability helps also to propagate changes among

different levels, namely concepts, requirements, specifications as well as take decisions or perform better impact studies. Traceability involves matrix-based methods (Krus, 2006) for mapping customer needs and requirements, design parameters and system behaviour. These methods allow re-evaluations in case of requirement changes (Luzeaux & Ruault et al., 2013). Besides traceability-oriented methods, it is possible to find requirements for traceability of SoS capabilities (Tekinerdogan, 2017).

A wildfire detecting SoS is used in the present work to illustrate some of the modelling and simulations considerations needed to develop traceable workflows. The scenario map, its environmental conditions, and the CS included must be allowed to change or evolve to study the SoS under different conditions. The outcome of the simulated scenarios is to be analysed for obtaining a set of requirements to constrain the design space of an electric Unmanned Aerial Vehicle (UAV). For the SoS explorations, the feasibility of Remote Component Environment (RCE) workflows will be evaluated with a focus on traceability. This paper aims to assess the convenience of using RCE's workflows in terms of traceability.

2. A Workflow-based Modelling Environment

The Remote Component Environment (RCE) is an open-source application developed by the German Aerospace Center (DLR) for design and simulation of systems (Boden, 2021). Serving as a flexible and scalable platform, it utilizes object-components, which represent a series of tasks to be executed in a specific order at predetermined intervals. RCE provides a graphical user interface for configuring the workflows, empowering users to define inputs, outputs, and the sequence of individual tasks. This structure ensures systematic and efficient experiments.

2.1. RCE Components

RCE features a range of components for different functions such as simulation, data processing and visualization. Central to these components available in RCE is the Design of Experiments (DoE) component, which allows users to define sets of input parameters with specific ranges and distributions, to set up and run experiments.

RCE also supports the incorporation of Python scripts as workflow components. This feature leverages Python's versatility, extending the range of tasks beyond the capabilities of the default RCE components. For example, conduct intricate calculations, manipulate data or files, or generate plots during the workflow runtime.

Additionally, RCE enables the integration of external tools as components in its workflows. This means that users can add their own software packages to RCE- for instance, by adding external tools for data acquisition, simulation, or analysis- and incorporate them into a workflow with built-in components for optimization.

2.2. System of Systems Experiments

Conducting SoS experiments for Aircraft Conceptual Design (ACD) is a complex application that requires the integration of multiple components for considering the CSs and their respective subsystems. The SoS experiments aim at optimizing requirements for a new aircraft, considering factors such as performance, safety, and cost. To achieve this, the DoE component within RCE is used to orchestrate and execute a series of experiments that delve into the compromises inherent between varying requirements and capabilities. The results of these experiments can then be analysed using Python scripts and external tools to identify the most promising design options. An example of an SoS workflow is shown in Fig. 1.

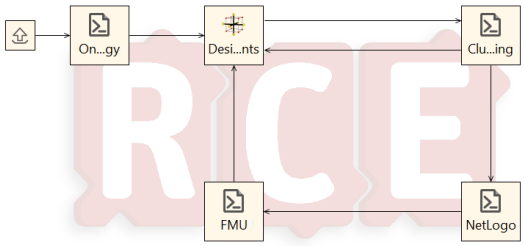


Figure 1: Structured workflow in RCE for SoS experiments.

3. SoS Workflow Construction

This section describes the different disciplines involved as block components in the assembly of the RCE workflow for wildfire detecting SoS exploration. Detection hinges on the subsystems nested within the CS. For this study, these subsystems encompass visual sensors with a resolution range. When a smoke plume is captured within the sensor's visual cone, detection is confirmed, marking the SoS's operation as successful and stopping the simulation.

3.1. Ontology Modelling

Ontologies serve as an instrument to formally encapsulate knowledge specific to a domain, including the concepts, relationships, and constraints that define it (Knöös Franzén, 2023). They provide a shared vocabulary and understanding for a group of people working on a common task, enabling more effective communication and reasoning about the domain. By defining a common ontology of the components, interfaces, and behaviours of the systems in the

study, it becomes easier for everyone involved to reason about how different systems interact with one another, and how changes done to one system will affect the others. Additionally, ontologies can facilitate the tracing of requirements, decisions, and outcomes across the SoS, which is important for understanding the impact of changes and making informed decisions (Lovaco, 2023). For visual clarity, Fig. 2 shows the ontology structure of a surveillance UAV. One ontology definition example (used for defining the present scenario) can be seen in Fig. 3.

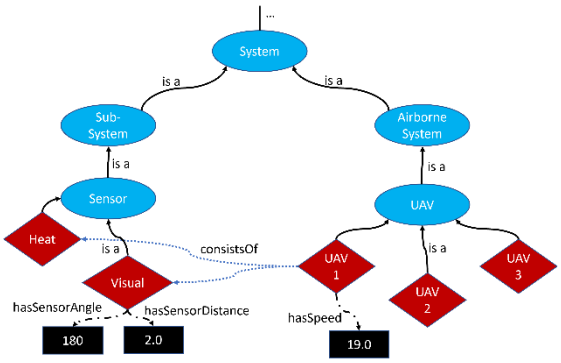


Figure 2: Partial ontology structure of a surveillance UAV.

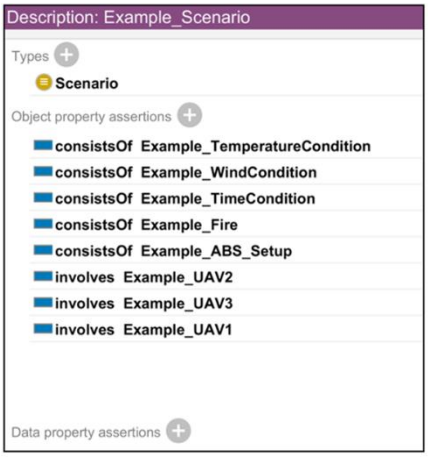


Figure 3: Ontology definitions of a surveillance UAV.

3.2. Clustering

Clustering can be described as the process of grouping a set of objects in such a way that objects in the same group (called a cluster) are in some manner more alike than those in different groups (clusters). Clustering is needed for navigation purposes since the flight paths will be defined afterwards based on the clusters to be visited.

The K-Means clustering algorithm is a popular method for clustering (Pedregosa, 2011). It partitions *N* data points into *K* clusters gauging their proximity to the centroids. Fig. 4 shows an example of the centres generated after clustering a given data set. However, determining the optimal number of clusters often poses a challenge. The method for determining the optimal number of clusters in K-Means clustering is explained in the section below.

3.2.1. Objective Function

The objective function of K-Means is to minimize the sum of squared distances between each data point and its assigned centroid:

$$\sum_{i=0}^{n} min_{\mu_j \in C} \left\| x_i - \mu_j \right\|^2 \qquad (1)$$

The time complexity of K-Means is O(KNT), where *N* is the number of data points and *T* the iteration number (Pedregosa, 2011).

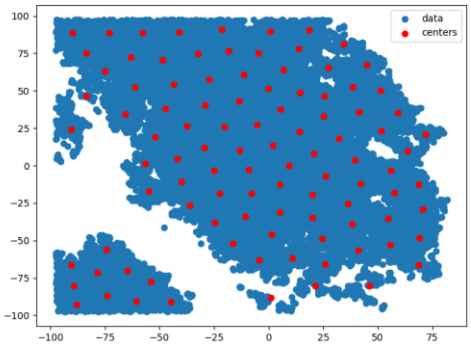


Figure 4: Centre points after clustering data.

A method for estimating the optimal number of clusters involves plotting the Within-Cluster Sum of Squares (WCSS) against the number of clusters and selecting the point where the WSS starts to level off (Pedregosa, 2011). The WCSS is defined as the sum of squared distances between each data point and its assigned centroid:

$$WCSS = \sum_{C_k}^{C_n} \left(\sum_{d_i in C_i}^{d_m} distance(d_i, C_k)^2 \right) \qquad (2)$$

Where *C* is the cluster centroid, *d* the data point and *n* the number of clusters.

The optimal number of clusters can be determined by inspecting the plot of WCSS as a function of the number of clusters, shown below in Fig. 5. Typically, the plot will show a steep decrease in WCSS as the number of clusters increases, followed by a levelling off. The "Elbow point" is the number of clusters at which the WCSS starts to level off. For the plot shown below, approximately 5 clusters yield a value of $0.2 \cdot 10^8$.

While visually perusing these plots is feasible, it might prove tedious and suboptimal, especially in large-scale simulations. In contrast, the use of a convergence analysis provides an automated and systematic approach. This allows for quicker identification of convergence behaviour and facilitates the exploration of design spaces in a more efficient and reliable manner.

A straightforward, yet more automation-friendly approach is to take the computed sum of squared distances and average it using the amount of data collected in that group. The result is compared with the squared value of a convenient parameter, which for the present case is the squared value of the distance that the sensors in the UAV can see.

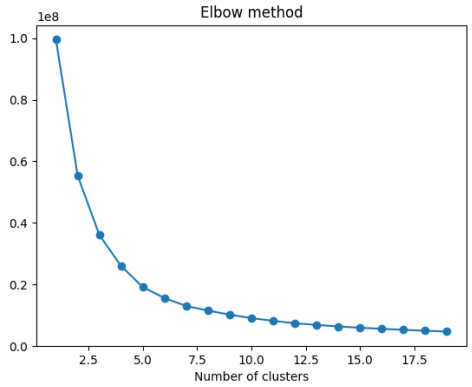


Figure 5: Evolution of the inertia depending on the number of clusters used.

3.3. Graph Theory for Finding the Shortest Navigation Path

The Traveling Salesman Problem (TSP) is a well-known combinatorial optimization problem that involves finding the shortest path to visit a set of points exactly once and return to the origin (Reinelt, 2003). One prevailing technique to tackle the TSP is by using graph theory (Euler, 1741), which provides an efficient and scalable solution (Hagberg, 2008). In this method, the TSP is modelled as a complete graph, where each point is represented by a node, and the edges represent the distances between them. The objective is to find the Hamiltonian cycle, which is a path that visits all nodes exactly once and returns to the starting node. Various algorithms, such as Christofides algorithm (Christofides, 1976), Simulated Annealing (Kirkpatrick, 1983), and Threshold Accepting (Dueck, 1990), can be used to find the shortest path. By using graph theory, the TSP can be solved with high accuracy, making it a useful tool for solving optimization problems in various fields. For the SoS use case, the centroids found in the previous step are clustered again depending on the number of CSs to be used, which for the present case is the number of UAVs. The TSP is solved for each aircraft to generate the navigation path for patrolling over an area. The Christofides algorithm is the chosen one for the present work, and it is used over a complete graph *G* to generate paths such as the one shown in Fig. 6. The initial node for the path is 0, but the nodes are not necessarily visited in the same order as they are numbered. The Christofides algorithm consists of the following steps (Goodrich & Tamassia et al., 2015):

1. A minimum spanning tree, *M*, is constructed for *G*.
2. Compute the set *W* with the vertices with odd degree from *M*. Form a new graph *H* with these vertices and the edges connecting them in *G*. Compute a minimum-weight matching *P* in this subgraph *H*.
3. Then *H* and *P* are combined into *G'* keeping repeated edges.
4. Find an Eulerian path *C*.
5. Finally convert *C* into a tour by skipping each vertex that has already been visited.

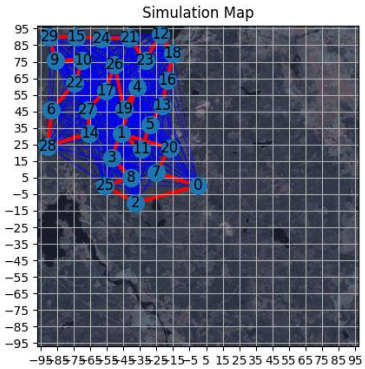


Figure 6: Hamiltonian cycle generated from a TSP solution.

3.4. Agent-Based Simulations for SoS Exploration

Agent-Based simulations (ABS) are a distinctive modelling and simulation technique that focuses on the behaviour of individual agents. Agents are entities with the ability to perceive their environment, make decisions, and act on their environment.

ABS boasts several advantages when compared to other modelling and simulation techniques. For one, ABS allows for the modelling of heterogeneity among agents, they can have different attributes, behaviours, and interactions, which makes the system more realistic (Lovaco, 2022). Moreover, ABS also allows to observe emergence, which is the phenomenon where the behaviour of the SoS is unequal to the sum of its individual parts. It must be noted that the resulting emergent behaviour should be judged as an increase or decrease of the SoS group value. Consequently, ABS is a promising approach to exploring the behaviour of interacting systems.

The exploration of complex SoS requires sophisticated tools for modelling and simulation. The tool used for this paper is NetLogo, an open-source ABS software with a drag-and-drop interface to create Agent-Based Models (ABM) (Tisue, 1999). Fig. 7 shows a NetLogo model interface created for firefighting SoS studies. NetLogo has several built-in capabilities for creating and manipulating agents, defining their behaviours, and visualizing simulation results.

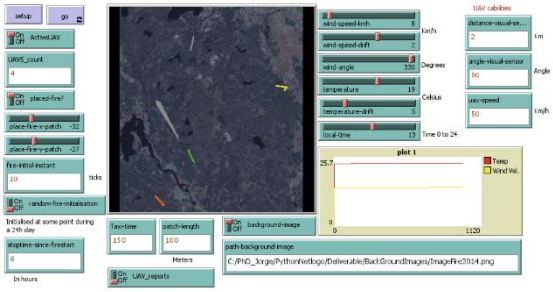


Figure 7: Agent-Based Model Interface.

3.5. Functional Mockup Units

The results extracted from the ABS can be analysed to find, for example, range and endurance requirements for UAVs. Once these specifications are distinctly outlined, cyber-physical UAV models with higher fidelity than the ones defined in the ABS are more suitable for ACD. These high-fidelity models can be made tool independent by using Functional Mockup Units (FMUs). Gaining traction in contemporary engineering and modelling sectors, FMUs are software modules that enshroud a specific functionality of a larger model of a system (Blochwitz, 2011). The main objective of FMUs is to enable model share and exchange between different tools and platforms in a standardized way. This means that models developed in one software environment can be easily integrated and used in another, without the need for custom integration code. FMUs are designed to be platform and language independent, they can be used across different operating systems and hardware architectures. This positions FMUs as a pivotal asset to share work and collaborate on interdisciplinary complex modelling and simulation projects (Braun, 2013).

4. SoS Case study and Results

This section presents the case study of a homogeneous forest-fire spotting SoS that consists exclusively of UAVs patrolling a given area to detect wildfires. Their respective flight trajectories, however, are derived uniquely by harnessing the K-Means clustering algorithm—which groups proximate trees—and subsequently deploying the TSP to delineate the Hamiltonian cycle for every UAV. Clustering is performed using the radius around each centroid determined by the visual sensor proficiency of the UAV. To evaluate different SoS configurations, the DoE uses a Latin Hypercube Sampling (LHS) method to generate different CS for the scenario. The UAV capabilities and scenario parameters not generated using LHS are extracted from the ontology, which is stored in an XML file. The simulations are executed in Netlogo with the wildfire initial geographic position being at a random point in the map. The fire detection success or failure is reported for each scenario. The flight missions for each UAV are simulated using a high-fidelity model FMU to evaluate the State of Charge (SoC) of the batteries and determine if the UAV configuration can achieve their mission (Krus, 2012). The workflow diagram of this case study was shown in Fig. 1, which from left to right starts by loading and reading the ontology; then the DoE component generates the different CS; the vegetation data is clustered afterwards to generate the UAVs flight paths; the NetLogo ABS executes the different SoS configurations; finally, the high-fidelity FMUs are executed to evaluate the performance of each aircraft concept. The workflow cycle is repeated as many times as initially defined in the DoE component. All the outputs forwarded at each step of the workflow are stored by RCE and accessible for the user once the experiment is completed, which is key for traceability.

4.1. Experiment Workflow

An ontology akin the represented in Fig. 2 is used to describe the scenario. The ontology XML file is read to extract the information needed for the SoS experiments. For this paper the values extracted from the ontology are the scaling factor, the fire detection time limit, position of the fire, wind velocity and wind direction. The number of UAVs used, and their capabilities are generated using the LHS and are catalogued in Tab. 1. This data is pivotal in tailoring the ABS, enabling a thorough examination of the ramifications stemming from diverse SoS layouts.

Table 1: LHS Generated Experiments.

ID	UAVS	Sensor Range [km]	Velocity [km/h]
1	6	1	88
2	7	2	106
3	9	1	153
4	4	2	190

Navigation routes are formulated in alignment with methodologies delineated in Sections 3.2 and 3.3. The paths depend on the sensor visual range and the quantity of UAVs defined. Fig. 4 shows the discerned centroids and Fig. 6 offers a glimpse into one such navigation route. The nondimensional ranges computed are shown in Tab. 2. The “Min Range” and “Max Range” values in the table represent respectively the minimum and maximum nondimensional distance flown by the UAVs in the SoS configuration. The scaling factor aids in converting these nondimensional distances into SI units.

Table 2: Nondimensional Ranges and Map Scaling Factor.

ID	Min Range	Max Range	Scaling Factor [m]
1	329	465	100
2	278	454	100
3	224	346	100
4	364	623	100

The ABS is performed for a map, as depicted in Fig. 8, where different surveillant agents can be seen. The map geography is generated importing elevation points into the model. The vegetation is dispersed randomly under certain topographic criteria, such as lake zones, resulting into the data distribution shown previously in Fig. 4. The geography and the fire location intend to relate to a real fire case (MSB, 2014). The arrow in the figure indicates the source of the smoke plume and the coloured lines represent a fraction of the flight paths taken by the aircraft systems.

Figure 8: Agent-Based simulation visualization.

The scenario was simulated using a quartet of different SoS configurations, each of which was executed ten times with wildfires initialized randomly, accumulating a total of 40 runs. Tab. 3 outlines the detection success rate and the averaged time required to spot the smoke plume for each SoS configuration.

Following the workflow sequence from Fig. 1, a high-fidelity aircraft model available in the cyber-physical high-performance simulation tool Hopsan (Krus, 2012) (which is similar to e.g., Modelica) was simulated through an FMU. This was done to ascertain the SoC of the battery, as well as the propeller diameter required to achieve the necessary velocity.

Table 3: Detection success and time.

ID	Success Rate [%]	Avg. Detection Time[min]
1	100	92
2	100	139
3	90	128
4	90	70

The necessary endurance, corresponding to the most extensive range for each UAV setup, is calculated and used to define the stop time for each FMU simulation. Tab. 4 summarizes the results extracted from these simulations: the battery SoC, the required diameter for the propeller, and the required endurance to complete just one loop of the predefined flight path.

Table 4: Aircraft FMU values.

ID	SoC [%]	Propeller Ø [cm]	Endurance [min]
1	73.0	28	32
2	78.2	28.5	26
3	86.5	31.5	14
4	78.1	34	20

5. Summary and Discussion

The workflow created ran all the 40 cases to study a firefighting SoS in 39 minutes real time. The obtained results can be used for constraining the design space to generate aircraft concepts, as well as discussing the trade-offs needed to achieve the capabilities of the desired aircraft concept. For a good trade-off analysis, a more extensive set of results is needed. But the procedure is exemplified in this section. First, Tab. 5 shows the collection of different CS capabilities.

Table 5: Constituent Systems Specifications.

CS	Sensor Range [km]	Velocity [km/h]	Range [km]	Prop. Ø [cm]	Endurance [min]
1	1	88	46.5	28	32
2	2	106	45.4	28.5	26
3	1	153	34.6	31.5	14
4	2	190	62.3	34	20

Then a comparison of SoS capabilities with their success rates (Tab. 4) helps in ranking requirements. This ranking can be subjective, based on the success rates, or on customer pre-defined criteria, culminating in a priority matrix. In Tab. 6, values 2, 1, or 0 denote higher, equal, or lower importance. Notably, below the diagonal, values swap between 0 and 2, while 1 stays constant. The trade-offs, discussed by rows are as following: Sensor Range (SR) holds higher significance than CS Velocity (V) since a higher speed does not increase the success;

comparison of SR with Range (R) or Endurance (E) needs more data for a decision hence it is assumed equal importance; SR, however, is seen as more crucial than Propeller Diameter (Ø) for fire detection.
The comparison is only exemplified here since it is beyond the current aim. It needs to be mentioned that, after identifying and discussing trade-offs to prioritize requirements, the ranking is eventually combined with methods such as the House of Quality (HoQ) or Quality Function Deployment (QFD) (Ulrich & Eppinger et al., 2016).

Table 6: Customer Requirement Priorities.

	SR [km]	V [km/h]	R [km]	Ø [cm]	E [min]
SR		2	1	2	1
V	0		0	2	1
R	1	2		2	2
Ø	0	0	0		1
E	1	1	0	1	

Fig. 9 shows an approximate curve representing the achieved cruise velocity as a function of the propeller diameter for a specified aircraft body geometry, which once they are obtained it is possible to find feasible regions for the design spaces.

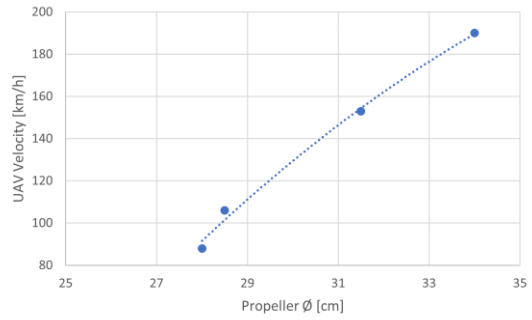


Figure 9: Achieved UAV cruise velocity as a function of the propeller diameter.

From the analysis, a notable topic is the aircraft's taxiing time. As indicated in Tab. 4, the aircraft can complete several missions along its flight path before needing to taxi, with the goal of preserving a SoC above 20% ideally. If this is not achieved, there might be a need to redesign the battery and aircraft to lessen their weight. This reduction can decrease both cruise speed and associated costs, potentially prompting another iteration of ACD.

For battery sizing, the drag forces over the mission can be extracted from the FMU results as well. Fig. 10 shows the flight altitude, and the aircraft drag over mission time. By knowing the altitude, it is possible to obtain the atmosphere properties. By understanding the forces that the aircraft needs to overcome to sustain the flight, it is possible to obtain the work and energy needed to fulfil the mission, and thus size a battery accordingly. Furthermore, studying the drag forces can reveal the parts of the geometry with the higher contribution to them and initiate a subsystem optimization process. These results can be obtained thanks to the use of a high-fidelity FMU in the workflow. It helps to explain, for example, the reduction over time of the drag forces observable in Fig. 10, which can be attributed to the decreasing energy in the battery that is translated into a reduced cruise velocity over the mission time.

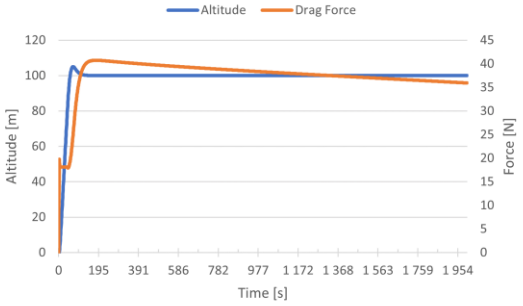


Figure 10: Aircraft altitude and drag forces values throughout time.

6. Future Work

The SoS studies for ACD can be expanded through the compatibility of RCE with XML files and CPACS files (Alder, 2020). CPACS files are designed for storing information related to aircraft concept generation. They can be used for encapsulating data pertinent to aircraft concept development, present a means to consolidate and disseminate expertise across diverse engineering domains, encompassing aerodynamics, propulsion, and structures.
RCE is equipped with dedicated features for the integration and utilization of other CPACS compatible tools in its workflows. One example of a tool designed to work seamlessly with CPACS files is TiGL (Siggel, 2019). TiGL is open-source and uses geometry libraries to visualize the 3D geometric representations of data extracted from CPACS. TiGL is also capable to export to several computer-aided design (CAD) formats.
Another promising future inquiry hinges on multidisciplinary optimization. The results of the SoS exploration studies can be analysed using HoQ and QFD methods for creating SoS value and system cost functions from the requirements. Subsequently analysing the balance between SoS value and CS cost to set the stage for a clearly delineated design space, marked by discernible requirement constraints, primed for preliminary concept evolution and optimization.

7. Conclusion

The research presented here using RCE workflows offers a robust method for tracing requirements and capabilities in SoS studies. A case study for aircraft concept generation to be used for wildfire detection was introduced to illustrate the created workflow. From parameters generated using LHS, navigation routes were generated from areas clustered based on the UAVs visual subsystems capabilities. The aircraft concepts and their routes were simulated to compare SoS mission success rates and constrain the design space of the aircraft concepts. Overall, this work illustrates how RCE workflows can be used for aircraft requirement generation by providing easy access to the results of every experiment allowing to sustain a high level of traceability.

Acknowledgment

The authors would like to acknowledge the Swedish Innovation agency (VINNOVA) for its financial support through the grant 2019-05371.

References

Mori, M., Ceccarelli, A., Lollini, P., Frömel, B., Brancati, F., & Bondavalli, A. (2018). Systems-of-systems modeling using a comprehensive viewpoint-based SysML profile. Journal of Software: Evolution and Process, 30(3), e1878.

Staack, I., Amadori, K., & Jouannet, C. (2019). A holistic engineering approach to aeronautical product development. The Aeronautical Journal, 123(1268), 1545-1560.

Dahmann, J. S., & Baldwin, K. J. (2008, April). Understanding the current state of US defense systems of systems and the implications for systems engineering. In 2008 2nd Annual IEEE Systems Conference (pp. 1-7). IEEE.

Maier, M. W. (2014). The role of modeling and simulation in system of systems development. Modeling and simulation support for system of systems engineering applications, 11-41.

Krus, P. (2006, September). Aircraft System Optimization and Analysis for Traceability in Design. In 11th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference (p. 7017).

Tekinerdogan, B., & Erata, F. (2017, April). Modeling traceability in system of systems. In Proceedings of the Symposium on Applied Computing (pp. 1799-1802).

Blochwitz T., Otter M., Arnold M., Bausch C., Clauß C., Elmqvist H., Junghanns A., Mauss J., Monteiro M., Neidhold T., Neumerkel D., Olsson H., Peetz J.-V., Wolf S. (2011): The Functional Mockup Interface for Tool independent Exchange of Simulation Models. 8th International Modelica Conference, Dresden 2011.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. the Journal of machine Learning research, 12, 2825-2830.

Hagberg, A., Swart, P., & S Chult, D. (2008). Exploring network structure, dynamics, and function using NetworkX (No. LA-UR-08-05495; LA-UR-08-5495). Los Alamos National Lab.(LANL), Los Alamos, NM (United States).

Lovaco, J., Staack, I., & Krus, P. (2022). Environmental Agent-Based Modelling For A Firefightingsystem Of Systems. In ICAS2022.

Lovaco, J. L., Franzén, L. K., & Krus, P. (2023, April). Agent-Based Simulation and Ontology Integration for System-of-System Exploration. In Proceedings of IDEAS 2022: Interdisciplinary Conference on Innovation, Design, Entrepreneurship, and Sustainable Systems (pp. 13-23). Cham: Springer International Publishing.

Axin, M., Braun, R., Dell'Amico, A., Eriksson, B., Nordin, P., Pettersson, K., ... & Krus, P. (2010). Next generation simulation software using transmission line elements. In Fluid Power and Motion Control, 15th-17th September, Bath, England, UK (pp. 265-276). Centre for Power Transmission and Motion Control.

Modelica Association. Modelica - A unified ObjectOriented Language for Physical Systems Modeling Language Specification - Version 3.2, 03 2010.

Krus, P., Braun, R., Nordin, P., & Eriksson, B. (2012). Aircraft system simulation for preliminary design. In 28th International Congress of the Aeronautical Sciences, Brisbane, Australia, 23-28 September, 2012 (pp. Art-nr). Optimage Ltd.

Braun, R., & Krus, P. (2013). Tool-independent distributed simulations using transmission line elements and the Functional Mock-up Interface. In 53rd SIMS conference on Simulation and Modelling, October 4-6, Reykjavik, Iceland.

Alder, M., Moerland, E., Jepsen, J., & Nagel, B. (2020). Recent advances in establishing a common language for aircraft design with CPACS.

Siggel, M., Kleinert, J., Stollenwerk, T., & Maierl, R. (2019). TiGL: an open source computational geometry library for parametric aircraft design. Mathematics in Computer Science, 13(3), 367-389.

Boden, B., Flink, J., Först, N., Mischke, R., Schaffert, K., Weinert, A., ... & Schreiber, A. (2021). RCE: an integration environment for engineering and science. SoftwareX, 15, 100759.

Myndigheten för samhällsskydd och beredskap (MSB). Skogsbranden i Västmanland 2014. MSB798, 2015. ISBN: 978-91-7383-527-5

Tisue, S., & Wilensky, U. (1999). Center for Connected Learning and Computer-Based Modeling Northwestern University, Evanston, Illinois. NetLogo: A Simple Environment for Modeling Complexity, Citeseer.

Christofides, N. (1976). Worst-case analysis of a new heuristic for the travelling salesman problem. Carnegie-Mellon Univ Pittsburgh Pa Management Sciences Research Group.

Dueck, G., & Scheuer, T. (1990). Threshold accepting: A general purpose optimization algorithm appearing superior to simulated annealing. Journal of computational physics, 90(1), 161-175.

Kirkpatrick, S., Gelatt Jr, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. science, 220(4598), 671-680.

Knöös Franzén, L. (2023). A System of Systems View in Early Product Development: An Ontology-Based Approach (Doctoral dissertation, Linköping University Electronic Press).

Euler, L. (1741). Solutio problematis ad geometriam situs pertinentis. Commentarii academiae scientiarum Petropolitanae, 128-140.

T. Reinelt, G. (2003). The traveling salesman: computational solutions for TSP applications (Vol. 840). Springer.

Goodrich, M. T., & Tamassia, R. (2015). Algorithm design and applications (Vol. 363). Hoboken: Wiley.

Luzeaux, D., & Ruault, J. R. (Eds.). (2013). Systems of systems. John Wiley & Sons.

Ulrich, K. T., & Eppinger, S. D. (2016). Product design and development. Boston: McGraw-Hill higher education.

Thermodynamics analysis of a novel compressed air energy storage (CAES) system combined with SOFC-MGT and using low grade waste heat as heat source

Chen Yang [a, *], Li Sun [b],

[a] *Chongqing University, Shapingba District, Chongqing 400044, China*
yxtyc@cqu.edu.cn

Abstract

As modern societies face increasing energy demands and a complex smart grid with multiple inputs of traditional and intermittent renewable energy power generation systems, the need for energy storage systems has become a general trend. Among these systems, compressed air energy storage (CAES) has received extensive attention due to its low cost and high efficiency. This study proposes a novel design framework for a hybrid energy system comprised of CAES system, gas turbine, and high-temperature solid oxide fuel cells, aiming for power generation and energy storage solutions. The overall model of the hybrid power generation system was constructed in Aspen Plus, and the mass balance, energy balance, and thermodynamic properties of the thermal system were simulated and analyzed. The results demonstrate that the hybrid system utilizes the functional complementarity of CAES and solid oxide fuel cells (SOFC), resulting in the cascade utilization of energy, flexible operation mode and increased efficiency. The overall round trip efficiency of the system is 63%, and the overall exergy efficiency is 67%, with a design net power output of 12.5 MW. Additionally, thermodynamic analysis shows that it is advisable to operate the system under higher compressor and turbine isentropic efficiencies, and optimal SOFC/MGT(Micro Gas Turbine) split air flow rates. The results of this article provide guidance for designing innovative hybrid systems and system optimization.

1. Introduction

For nearly a century, the use of traditional fossil fuels has provided a stable and huge contribution to the progress of human civilization. But with a growing population and improved living standards, energy demand is expected to increase dramatically. According to statistics, it is estimated that by 2100, the global population will reach 10 billion. Meanwhile,emissions will increase correspondingly if immediate action is not taken. Looking at it today, the world emits about 51 billion tons of greenhouse gases into the atmosphere every year. In order to prevent climate disasters, countries around the world have established relevant policies and measures to reduce carbon emissions and strive to reduce carbon emissions to zero by 2050(IPCC,2022). This indicates that the most imminent and important task within the next three decades is to improve the efficiency of the existing energy systems, and to increase the share of renewable energy in the energy sector.

However, the utilization of renewable energy sources is accompanied by uncertainties and instabilities due to environmental factors such as wind and solar radiation. The integration of large-scale renewable energy into the power grid can have significant impacts on grid security and compromise the quality of electrical energy. This poses substantial challenges for maintaining power balance and ensuring stability control in the grid. To address these challenges, energy storage technologies have emerged as essential solutions. Energy storage technologies offer the capability for large and medium-scale energy storage and rapid response. When integrated into comprehensive energy systems that encompass conventional fossil fuel power generation and renewable energy sources, energy storage technologies play a pivotal role in maintaining power and energy balance across various operating conditions. Consequently, energy storage holds crucial implications for enhancing grid security, facilitating economic operation, and maximizing the utilization of renewable energy sources (Bazdar E et.al,2022).

Among various energy storage technologies, Advanced Adiabatic Compressed Air Energy Storage (AA-CAES) technology has gained significant attention from researchers in the past decade due to its advantages of large-scale energy storage, high energy efficiency, and zero emissions (Zhang, W.2020). However, the operation of AA-CAES requires the storage of compressed heat during the energy storage period the utilization of compressed heat to heat the air during the energy

release phase (G. Grazzini and A. Milazzo,2008). This characteristic imposes high requirements on the capacity, efficiency, and control of the heat exchangers and thermal storage systems. Therefore, in order to improve the overall efficiency of Compressed Air Energy Storage (CAES) and reduce emissions, one potential solution is to explore the utilization of other thermal sources within the comprehensive energy system to achieve better energy grade utilization (Javier, M,2019). In addition, the integration of different thermal sources within the comprehensive energy system allows for greater flexibility and versatility in energy generation and utilization. It enables the system to adapt to varying energy demands and optimize the use of available resources based on specific conditions or requirements. (Li Y, et al.2019).

In order to evaluate the potential of Compressed Air Energy Storage (CAES) in enhancing renewable energy integration, researchers have developed a comprehensive approach by merging a fixed efficiency model and a detailed thermoelectric model of CAES with unit commitment and economic dispatch algorithms (Nikolakakis, T, et al.2017).

SOFC can be well applied to large-scale distributed power generation systems of hundreds of megawatts. Typically, SOFC is operating at a high temperature range of 800-1000℃ with capabilities of high energy conversion rate, high current density and power density, wide fuel applicability, good electrode economy, modular assembly, and low emission. In particular, SOFC has strong reaction kinetics and is hardly poisoned by impurities such as CO and its by-products, so it has received widespread attention from researchers all over the world. However, commercially available solid oxide fuel cells (SOFCs) can only operate safely in steady-state mode, and therefore they need to be integrated with an energy storage system for use in load-following applications. Nease and Adams (2014a) proposed a coal-fueled SOFC plant integrated with compressed air energy storage (CAES) which had no CO2 emissions. And then they referred a two-stage rolling horizon optimization (RHO) framework is used to optimize a SOFC/CAES integrated power plant, achieving optimal year-round peaking power with zero emissions and significantly improving load-following performance by up to 90% (J. Nease and T.A. Adams II,2013) (Nease, J, et al.2016).

This study aims to figure out the interaction between a SOFC/MGT system and an AA-CAES system by considering the energy and exergy balance. The study also investigates the impact of different input parameters on novel system performance. The contributions of this work are summarized as below.

First, a new hybrid system is proposed in this paper, namely the AA-CAES system integrated with the SOFC/MGT system. Although there have been scholars who have done similar integrations in the past, but it is only a simple superposition, the system integration in this article is more thorough, and the combination between subsystems is more precise. This is mainly reflected in the thermal coupling between subsystems in this paper is designed based on the characteristics of each subsystem itself. That is to say, this paper not only realizes the cascade utilization of energy according to the difference of temperature, but also considers the safety and actual operation requirements of the subsystem. Moreover, unlike other similar systems, the CAES system in this paper is upstream of SOFC/MGT, addressing the pain points of the control problems caused by the interconnection of compressors and turbines and deep coupling with SOFC in traditional SOFC/MGT systems.

Secondly, when considering round trip efficiency and exergy efficiency in this paper, different analysis methods are proposed because of the time-segmented characteristics of energy storage systems. It considers not only the performance of the system under different operating modes, but also the intrinsic characteristics of different subsystems.

Finally, the impact of the study of key parameters on system performance can provide the goal of system optimization and also enhance the actual engineering value of the entire system.

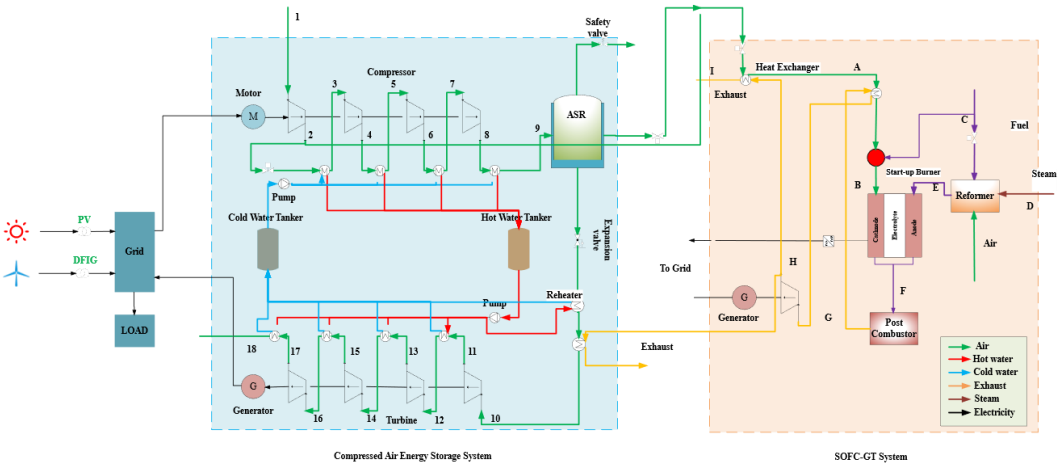


Figure 1. Overview of the CAES-SOFC/MGT Hybrid system

2. Methodology

It can be seen the working process of the integrated system used for this work from

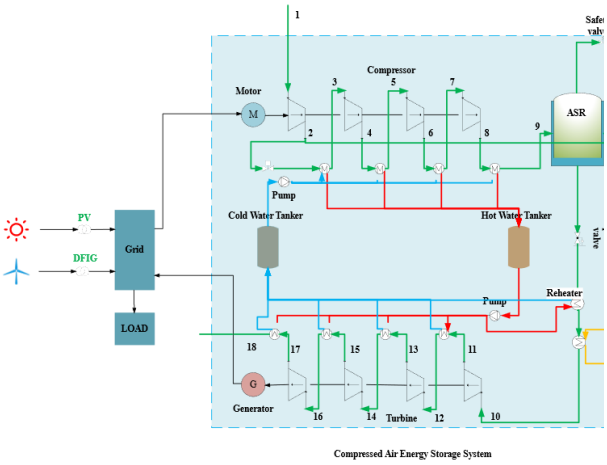


Figure *1*. Firstly, it is assumed that the SOFC-MGT system is always working as the base load. The hybrid power generation system combines compressed air energy storage, fuel cells, and a gas turbine. The system includes an air compressor unit, energy release turbine unit, heat storage tanks, gas storage tank, generator, motor, regenerator, and fuel cell power generation system. The motor drives the air compressor unit to compress air when surplus power is available. The compressed air is stored and released through the turbine group, connected to the generator. The fuel cell system is connected to the grid and includes a start-up burner, reactor, fuel cell body, afterburner, and post-combustion turbine. The high-pressure air from the compressor is heated by the microturbine exhaust for waste heat recovery before entering the fuel cell. In case of peak load, the compressors use the surplus electricity from renewable energy generation and compress the air into the storage tank in the charge mode. The stage inter cooling heat exchanges pump the heat into a hot water tank. Meanwhile, the first compressor of the charge chain works for SOFC. The compressed air without cooling supplies the cathode. Besides, before flowing into the fuel cell, the high-pressure air is heated by exhaust gas from the gas turbine, which can make the utmost of the wasted heat. The fuel cell generates the power to grid and the high pressure, temperature by-product with remain fuel flows into the turbine to generate power too. In the discharge mode, the air turbines are also working in four stages. The gases coming into the air turbines are heated up by the transfer fluid both in hot water tank and gas turbine. The cooling water goes to the cold water tank driven by the pump system.

Since the operation of the energy storage system has a time difference, in order to better analyze the operation of each component in each stage, this paper divides the operation of the entire system into five states, which can be seen from Figure *2*.

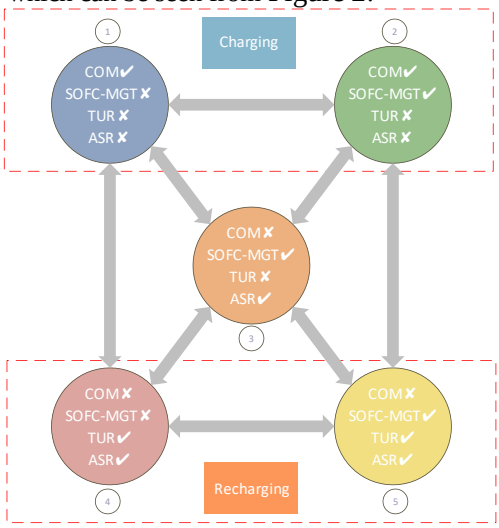


Figure 2. Working states of the CAES-SOFC/MGT Hybrid system.

In Figure 2, mode 1 to 5 represent the operating states of each subsystem, where the checkmark symbol indicates that the subsystem is running, and the cross symbol indicates that the subsystem is not running. Mode 2 and 5 represent the typical operating states of the system, while mode 1, 3, and 4 are used only in extreme conditions or for testing the system performance.

When the grid load is below the preset energy storage load, the grid functions as a motor to drive the air compressor unit, which compresses and stores the air in the gas storage device. During compression, the cold water from the tank's cold water tank flows into the heat exchanger to cool the compressed air, and then flows into the hot water tank.

Mode 1 is a test mode used to detect the working mode of the compression chain when it runs alone. At this mode, the air storage reservoir is in the energy storage stage, while other subsystems are inactive.

Mode 2 is the normal energy storage mode of the system. At this mode, excess power from the grid drives the compressor to produce compressed air. Some of the compressed air is stored in the air tank, while the rest is used as the cathode inlet air of the fuel cell, serving as the SOFC-MGT air source.

Mode 3 is a test mode and also the minimum output power condition of the system, which is used to test the working mode of the SOFC-MGT system when it is running alone. In this mode, the air tank serves as the air source of the SOFC-MGT system, and the compressed air is used as the cathode inlet of the fuel cell. When the air tank is fully charged, the SOFC-MGT system can provide compressed air to the entire system for hours or even days, depending on its capacity. Therefore, this mode can be used as a reference for black start situations, where the energy storage system provides users with the minimum necessary output power when the external power grid is paralyzed.

Mode 4 is designed as a test mode to evaluate the energy release chain's working status when it operates independently. In this mode, the compressed air from the air tank is supplied to the air turbine, and the hot water in the hot water tank and the regenerator between each stage heat the compressed air. This process allows the compression heat in the compression stage to be fully utilized.

Mode 5 represents the normal energy release mode and the maximum output power of the system. In this mode, the compressed air in the air tank is split into two streams, of which one stream directly drives the air turbine to generate electricity, while the other stream serves as the anode inlet air for the SOFC-MGT system. Unlike mode 4, the high-temperature exhaust gas at the outlet of the SOFC-MGT system heats the inlet air of the first-stage air turbine, allowing for the full utilization of the waste heat from the exhaust gas.

The following assumptions are made to simplify the whole system modeling.

(1) The composition of air in CAES system and SOFC cathode inlet consists of 79% N_2 and 21% O_2 (mass fraction).
(2) The heat and pressure loss in the pipes connecting all the components can be negligible.
(3) All the kinetic and potential effects are ignored.
(4) The storage tank is adiabatic during the charging and discharging process.
(5) The water gas reaction is in a state of chemical equilibrium.
(6) The effect of radiation heat transfer is not considered.
(7) The system operates under steady state conditions.
(8) The isentropic efficiency of each compressor and turbine is fixed.

Basic geometric and operating parameters of this hybrid system are shown in Table 1 and

System parameters	Value
Ambient pressure / bar	1.000
Ambient temperature / k	298.0
Maximum pressure of ASR /bar	100
Minimum pressure of ASR /bar	70
Volume of the air tank /m^3	6000
Hot water tank temperature / k	403.0
Hot water tank pressure / bar	4.000
Cold water tank temperature / k	298.0
Cold water tank pressure / bar	1.000
Generator efficiency	0.98
Inlet pressure / bar	70.00

Table *2*. The SOFC geometric parameters were taken from (Huang S, et al.2022):

Table 1. Basic operating parameters for CAES

System parameters	Value
Ambient pressure / bar	1.000
Ambient temperature / k	298.0
Maximum pressure of ASR /bar	100
Minimum pressure of ASR /bar	70
Volume of the air tank /m^3	6000
Hot water tank temperature / k	403.0
Hot water tank pressure / bar	4.000
Cold water tank temperature / k	298.0
Cold water tank pressure / bar	1.000
Generator efficiency	0.98
Inlet pressure / bar	70.00

Table 2.Operating parameters for SOFC/GT hybrid system

Operating parameters	Value
Total cell area	200.0×200.0 mm
Anode thickness	0.5 mm
Electrolyte thickness	0.008 mm
Cathode thickness	0.05 mm
Number of channels	50
Channel size	2.0×2.0 mm^2
Cathode inlet pressure	3.9 bar
Anode inlet pressure	3.9 bar
Average SOFC temperature	939 °C
Current density	0.676 A/cm2
Fuel utilization factor	0.62
Reformer inlet composition (molar fraction)	33% CH_4, 67% H_2O
Turbine isentropic efficiency	86%
Turbine mechanic efficiency	98%

2.1 Compressor

There are four air compressors in the system connected in series, with interstage coolers between each compressor, and an aftercooler after the fourth air compressor. The outlet pressure of the air compressor can be calculated by the following formula (Hartmann N, et al. 2012):

$$P_{out,c} = \pi_c P_{in,c} \tag{1}$$

where $P_{in,c}$ is the inlet pressure of the compressor, π_c is the compression ratio of the air compressor, and the inlet pressure of the first stage compressor is atmospheric pressure. The outlet temperature of the compressor is:

$$T_{out,c} = T_{\mathrm{in},c}[1+(\pi_c^{(\mathrm{k}-1)/\mathrm{k}}-1)/\lambda_c] \tag{2}$$

where $T_{\mathrm{in},c}$ is the inlet temperature and λ_c is the isentropic efficiency of the compressor. k is the specific heat ratio. The power consumed by the compressor is:

$$\begin{aligned}\dot{\mathrm{W}}_{com} &= \dot{m}_{c,in}(\mathrm{h}_{c,out}-\mathrm{h}_{c,\mathrm{in}}) \\ &= C_{\mathrm{p}}\dot{m}_{c,in}T_{\mathrm{in},c}\frac{1}{\eta_c}(\pi_c^{(k-1/k)}-1)\end{aligned} \tag{3}$$

where $\dot{m}_{c,in}$ is the inlet air mass flow rate of the air compressor in the energy storage stage, $\mathrm{h}_{c,out}$ and $\mathrm{h}_{c,\mathrm{in}}$ are the specific enthalpy of the outlet and inlet air, respectively.

2.2 Heat exchanger

The heat of compression is transferred to the heat storage device through the interstage heat exchanger, and a counter-flow heat exchanger is used in this paper. The water temperature is approximately equal to ambient temperature before entering the interstage cooler. From the law of conservation of energy （Guo C, et al. 2017）.

$$\begin{aligned}&\dot{m}_{c,\mathrm{air}}(\mathrm{h}_{\mathrm{inter},air,in}-\mathrm{h}_{\mathrm{inter},air,\mathrm{out}}) \\ &= \dot{m}_{c,\mathrm{water}}(\mathrm{h}_{\mathrm{inter},\mathrm{water},\mathrm{out}}-\mathrm{h}_{\mathrm{inter},\mathrm{water},\mathrm{in}}) = \mu_{\mathrm{inter}}A_{\mathrm{inter}}\Delta T_{\mathrm{inter}}\end{aligned} \tag{4}$$

Where $\mathrm{h}_{\mathrm{inter},air,in}$ and $\mathrm{h}_{\mathrm{inter},air,\mathrm{out}}$ are the air specific enthalpy at the inlet and outlet of the interstage cooler, $\mathrm{h}_{\mathrm{inter},\mathrm{water},\mathrm{out}}$ and $\mathrm{h}_{\mathrm{inter},\mathrm{water},\mathrm{in}}$ are the specific enthalpy of the water in the heat storage system at the inlet and outlet of the interstage cooler, μ_{inter} is the thermal conductivity of the heat exchanger, and A_{inter} is the heat transfer area, $\Delta T_{\mathrm{inter}}$ is the logarithmic mean temperature difference between air and water.

There is a regenerator in front of each turbine, which is the same as the energy storage part, which is a counter-flow heat exchanger. In the regenerator, the hot water stored in the hot water tank is heated to the air flowing out of the air storage tank, and then in the cold water tank was cooled to room temperature.

$$\dot{m}_{c,\mathrm{air}}(\mathrm{h}_{\mathrm{re},air,out}-\mathrm{h}_{\mathrm{re},air,\mathrm{in}}) = \dot{m}_{c,\mathrm{water}}(\mathrm{h}_{\mathrm{re},\mathrm{water},\mathrm{in}}-\mathrm{h}_{\mathrm{re},\mathrm{water},\mathrm{out}}) \tag{5}$$

where $\mathrm{h}_{\mathrm{re},air,\mathrm{in}}$. and $\mathrm{h}_{\mathrm{re},air,out}$ are the specific enthalpy of air at the inlet and outlet of the regenerator, and

$h_{re,water,in}$ and $h_{re,water,out}$ are the specific enthalpy of water at the inlet and outlet of the regenerator.

2.3 Turbine

In this paper, four turbines are connected in series. The outlet pressure of the turbine can be calculated by the following formula (Hartmann N, et al. 2012)(Guo H, et al.2019):

$$P_{\mathrm{in,t}} = \pi_t P_{\mathrm{out,t}} \tag{6}$$

where $P_{\mathrm{in,t}}$ is the inlet pressure of the turbine, π_t is the expansion ratio of the turbine, and the inlet pressure of the first stage turbine is the outlet pressure of the gas storage tank. The outlet temperature of the turbine is:

$$T_{out,\mathrm{t}} = T_{\mathrm{in,t}}[1-\lambda_t(1-\pi_t^{(\mathrm{k}-1)/\mathrm{k}})] \tag{7}$$

where $T_{\mathrm{in,t}}$ is the inlet temperature of the turbine, and λ_t is the isentropic efficiency of the turbine. *k* is the specific heat ratio. Turbine work can be calculated by the following formula:

$$\begin{aligned}\dot{\mathrm{W}}_t &= \dot{m}_{\mathrm{t,air}}(\mathrm{h}_{\mathrm{t,in}} - \mathrm{h}_{c,\mathrm{out}}) \\ &= C_{\mathrm{p}}\dot{m}_{\mathrm{t,air}}T_{\mathrm{in,t}}\lambda_{\mathrm{t}}(\pi_t^{(k-1/k)}-1)\end{aligned} \tag{8}$$

2.4 Air storage tank

After flowing out of the interstage cooler, the hot water absorbing the heat of compression will be stored in the hot water tank.

In this paper, the gas storage device adopts the gas storage tank. In order to study the performance of the gas storage tank, according to the law of mass conservation and energy conservation (Wu S,et al.2019):

$$\begin{aligned}&Q_{\tan k} = \dot{m}_{c,in}h_{c,in} - \dot{m}_{t,out}h_{t,out} \\ &-\dot{m}_{\mathrm{split},out}h_{\mathrm{split},out} - Q_{\mathrm{loss}}\end{aligned} \tag{9}$$

$$m_{\tan \mathrm{k}} = \dot{m}_{\mathrm{c},in}\mathrm{t}_c - \dot{m}_{t,out}t_t - \dot{m}_{split,out}t_t \tag{10}$$

where $Q_{\tan k}$ and Q_{loss} is the rate of transferred heat, $\dot{m}$ is the air flow rate, t_c and t_t is the charge time and discharge time, and the subscript split represents air split from tank in the mode 5.

2.5 SOFC/MGT

The SOFC model used the lumped parameter method to calculate the local current density, distribution temperature distribution, Nernst potential, and electrochemical loss of the whole fuel cell. The chemical kinetic model was used to calculate the water gas shift reaction, steam methane reforming reaction, and electrochemical reaction. The SOFC model was created in ACM and imported into Aspen Plus TM. The current density can be calculated based on the feed flow rate of each component of the inlet fuel and the set fuel utilization rate. The operating conditions of the fuel cell are modeled in terms of inlet and outlet average temperatures, pressures, and stream composition. The cell voltage is calculated by Bulter-Volmer formula, and the power output of a fuel cell can be calculated based on the total reaction area of the cell and the total number of cells. Thomas Paul Smith et al. published a detailed description of the model and validation work. The electrochemical loss and cell voltage calculations are expressed by equations as follows (Liese E A, et al. 2006).

$$V_{cell} = V_{Nernst} - \eta_{act} - \eta_{con} - \eta_{ohm} \tag{11}$$

$$V_{Nernst} = E^0 + \frac{RT}{2F}\ln[\frac{P_{H_2}P_{O_2}^{0.5}}{P_{H_2O}}] \tag{12}$$

$$E^0 = 1.2877 - 0.0002904T_{cell,bulk} \tag{13}$$

$$\eta_{\mathrm{con}} = \frac{RT}{2F}\left(\ln(\frac{\chi_{H_2,bulk}\chi_{H_2O,TPB}}{\chi_{H_2O,bulk}\chi_{H_2,TPB}})\right) + \frac{1}{2}\ln(\frac{\chi_{H_2,bulk}}{\chi_{O_2,TPB}}) \tag{14}$$

$$\eta_{act} = \frac{RT}{2\alpha F}\sinh^{-1}\left(\frac{i}{2i_o}\right) \tag{15}$$

$$\eta_{ohm} = \left[\begin{array}{l}6.78E-10\exp(\frac{10.3E3}{T_{cell,bulk}}) + \\ 0.23174\exp(-0.0115T_{cell,bulk})\end{array}\right] * i \tag{16}$$

This paper uses the ACM model in the Aspen Plus software model library to establish the SOFC battery model. Before flowing into the cathode, a part of the air is split into reformer, combusted with part of the fuel to make the reformer an autothermal reformer and ensure the temperature of the fuel at the anode inlet. The remaining air is discharged after exchanging heat with SOFC's outlet flue gas in the heat exchanger, while ensuring the inlet temperature of the after burner. At the anode, the natural gas undergoes a reforming reaction; the produced hydrogen reacts electrochemically with the air from the cathode to generate electricity.

The input data for natural gas, air volume, and water vapor required for the stack model were determined by the calculator model of the ASPEN plus. Assuming that the cells in the stack are connected in parallel, the voltage of each battery is the same. The current of the fuel cell stack is equal to the current per cell multiplied by the number of cells, which calculated in the SOFC ACM model.

2.6 Round trip efficiency

When dealing with a hybrid power system that includes an energy storage system, it can be challenging to determine a precise efficiency value, as the charging and discharging periods operating at divergent times. In such cases, the system's round-trip efficiency (RTE) is used as a metric. RTE is ratio of total electrical energy output to the total fuel and electrical energy input for a full charge/discharge cycle and can be calculated using the following formula:

$$RTE = \frac{\int_T W_{OUT}}{\int_T W_{IN}} = \frac{\int_{T_{dc}} W_t + \int_{T_{all}} (W_{MGT} + W_{SOFC}) + Q_{HL}}{\int_{T_{ch}} W_c + \int_{T_{all}} (\dot{m}_{CH_4} \cdot LHV_{CH_4})} \quad (17)$$

Where W_{OUT} is the electricity out , W_{IN} is the energy input, W_t is the electricity generation for expanders, W_{MGT} is the electricity generation for turbine, Q_{HL} is the excess heat load in hot water tank for end user. W_{SOFC} is the electricity generation for SOFC, W_{c} is the electricity consumption of compressors, $\dot{m}_{CH_4}$ is the mass flow of inlet CH_4 , LHV_{CH_4} is the lower heating value of CH_4 , T_{dc} is the discharge time, T_{ch} is the charge time, T_{all} is the one-cycle time.

2.7 Total exergy efficiency

Similar to the definition of system cycle efficiency, total exergy efficiency (TEE) is defined as the total net exergy output E_{xout} in discharging process to total net exergy input E_{xin} in the charging process. As the heating exergy is not considered in the reference, TEE can be presented as:

$$TEE = \frac{E_{xout}}{E_{xin}} = \frac{E_{xTB} + E_{xSOFC} + E_{xMGT}}{E_{xCOM} + E_{xCH4}} \quad (18)$$

where E_{xTB} is the total exergy of expanders, E_{xCOM} is the total exergy of compressors, E_{xSOFC} is the total exergy of SOFC, E_{xMGT} is the total exergy of turbine and E_{xCH4} is the total exergy of fuel.

3. Results and discussion

3.1 Influence of outlet pressure of expansion valve

In discharging process, the air from the ASR passes through the expansion valve so that a specific pressure will be set, which makes the turbines operate at the steady and high efficiency conditions. The changing of the required pressure has a direct influence on the turbines' performance. The power output of the turbine and the mass flow rate of air increase with increasing pressure behind the expansion valve. However, the operating range of the turbine narrows, and the discharge time decreases sharply. The 0-D system model lacks a control mechanism, resulting in changes in mass flow rate in response to pressure variations. As shown in Figure 3, modes 4 and mode 5's efficiencies rise with increasing post-valve pressure. The throttle valve has no impact on the operation of the compression chain and SOFC/MGT subsystems. Hence the efficiencies of modes 1, 2, and 3 remain unchanged. As pressure behind the valve rises gradually, the system RTE drops from 65.9% to 49.6%. The reason is that although the turbine does more work per unit of time, but the release time for energy decreases, which reduces the working range of the air storage tank, ultimately decreasing the total work done by the turbine.

Figure 4 shows the effect of expansion valve outlet pressure on the total exergy efficiency. In the figure, nECOM represents the exergy efficiency of the compressor, nETUR represents the exergy efficiency of the expander, nESOFC-MGT represents the exergy efficiency of SOFC-MGT, and EXALL represents the exergy efficiency of the system total exergy efficiency. As seen from Figure 4, the TEE of the compressor chain and SOFC/GT subsystem keep constant as well. But turbine chain subsystem's TEE increases with the varying expansion valve outlet pressure from 60 bar to 80 bar. It is evident that when the inlet pressure is increased, each turbine performs more work, leading to an increase in exergy efficiency. While the flow rate also rises, requiring more hot water heat and leading to greater exergy loss in the turbine chain's regenerator, these effects are relatively minor in comparison to the gain in exergy efficiency resulting from the increased work done by the turbines. However, TEE exhibits a clear downward trend. This is due to the significant reduction in the energy release time, despite the increase in work done per unit time by the turbine. Compared to the aforementioned effects, the energy discharge time has a greater impact on the total output exergy. In conclusion, the highest achievable values of the overall system's round trip efficiency and exergy efficiency are obtained when the heat load output to the user approaches zero, while the turbine maximizes the use of the compression heat and SOFC/MGT exhaust waste heat.

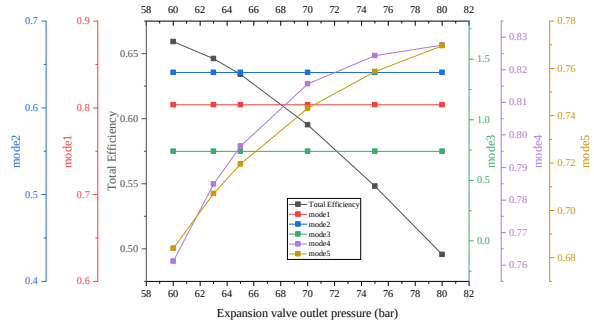


Figure 3. Effect of expansion valve outlet pressure on RTE

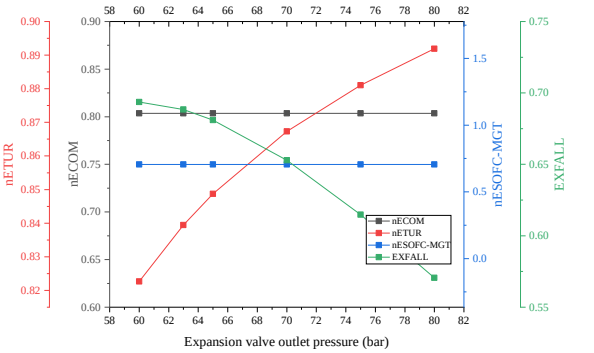


Figure 4. Effect of expansion valve outlet pressure on TEE

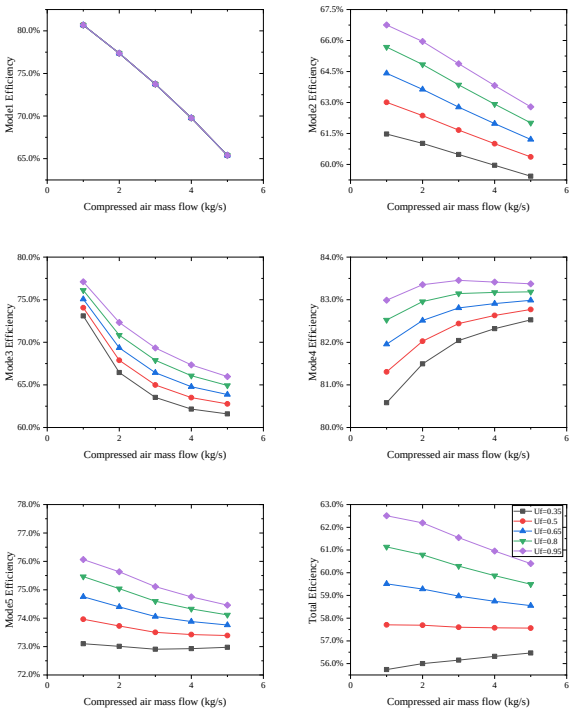


Figure 5. Energy flow diagram of CAES-SOFC/MGT Hybrid system

3.2 Influence of fuel utilization factor and air mass flow to SOFC/MGT system

Figure 5 presents a comprehensive analysis of the impact of varying the fuel utilization factor and SOFC/MGT air inlet mass flow on the efficiency of the hybrid system in different operating modes. Specifically, the figure demonstrates the effect of changing the aforementioned parameters on the overall efficiency of the hybrid system as well as the efficiencies of the individual subsystems in different operating modes.

As illustrated in Figure 5(a), the compressed air flow rate is the main factor affecting mode 1, which refers to the compression chain of the CAES system. When the compressed air flow rate diverted to the SOFC/MGT system increases, the flow rate of the compressor decreases. Consequently, the power consumption of stages 2-4 of the compressor decreases, along with a reduction in the compression heat. However, this also leads to a significant decrease in the energy stored per unit time, resulting in a reduction in the efficiency of mode 1. Because the efficiency reduction of the compressor is so large that even though the efficiency of the SOFC/MGT system has increased caused by the Uf(fuel utilization) changing, the efficiency in mode 2 is still reduced, but a higher Uf will result in higher efficiency, as shown in Figure 5(b). The efficiency of mode 3 is directly affected by the variation in Uf and the SOFC/MGT inlet flow rate. A high fuel utilization rate can reduce the fuel cell efficiency, and a large anode air flow can lower the average temperature of the fuel cell. However, since the compressed air in mode 3 mainly comes from the CAES subsystem's compressor, high fuel utilization rate leads to higher efficiency in mode 3. On the other hand, an increase in SOFC anode air reduces the heat output from the SOFC/MGT subsystem to the downstream, resulting in a decrease in efficiency. As shown in Figure 5(d), the increase in system efficiency is observed only at the energy release end of the CAES system. This is because the work done by the turbine per unit time remains unchanged, but the heat output to the turbine decreases. Combining the turbines with the SOFC/MGT, it can be seen that the efficiency reduction of the SOFC/MGT subsystem has a greater impact, so the efficiency under mode 5 is reduced.

Overall, the hybrid system's efficiency increases as fuel utilization increases, considering the changes in energy storage and release times due to variations in air flow. When fuel utilization is low, the SOFC/MGT subsystem's air flow has a positive effect on RTE. However, at high fuel utilization rates, the SOFC/MGT subsystem's air flow has a negative effect on RTE.

Figure 6 depicts the exergy efficiency of each subsystem and the overall system as a function of Uf

and the inlet air flow rate to the SOFC/MGT subsystem, while Figure 7 shows the variation of exergy loss of each main component and its proportion in the system under different combinations of Uf and inlet air flow rate. In these figures, Uf varies between 0.35 and 0.95, and the inlet air flow rate changes from 1kg/s to 5kg/s. The performance of the SOFC is strongly affected by Uf, which, in turn, impacts the performance of downstream components, such as the afterburner, MGT, and turbine chain. As shown in Figure 6, the exergy efficiency of the SOFC/MGT subsystem increases with Uf since a higher fuel utilization rate results in higher hydrogen reacting in the SOFC and, consequently, higher electric energy output. Additionally, the outlet temperature of the fuel cell is also higher, which reduces the combustible gas entering the combustion chamber downstream, leading to a reduction in the exergy loss of the combustion chamber. Consequently, the exergy loss of the MGT also decreases, and the degree of decrease has little effect on the decrease of the MGT's output work compared to the increase of the output work brought by the increase in the turbine inlet flow. The highest exergy efficiency of the solid SOFC/MGT subsystem is achieved when Uf is 95% and the inlet air flow rate is 5kg/s.

When the air flow to the SOFC/MGT subsystem increases, the flow rate of the 2-4 stage compressors in the compression chain of the CAES system decreases, and the power consumption also decreases, leading to a significant reduction in the compression heat of the entire compression chain. Furthermore, the exergy loss of the interstage cooler is reduced. As a result, the total exergy efficiency of the compression chain increases with an increase in the air flow diverted to the SOFC/MGT system. However, after a complete cycle, the split air flow rate reaches 5kg/s, and the difference between the heat of compression and the heat required for expansion of the entire system is negative, indicating that the system can no longer operate completely without additional external heat supply.

For the energy release chain of CAES, the input work of the turbine remains unchanged since the turbine inlet temperature and air flow rate remain constant. However, due to the reduction of the exhaust gas temperature of the SOFC/MGT system, the exergy loss of the heat exchanger between the compressed gas tank and the turbine chain is reduced, and the exergy efficiency of the solid-release energy chain slightly increases.

Based on the analysis above, combined with Figure 7, it can be concluded that for the entire hybrid system, the compression chain exergy loss of CAES is more sensitive to air flow changes, while SOFC/MGT has little effect on the exergy loss of the entire system. Therefore, the highest total exergy efficiency is 67.24%, which increases with the increase of Uf and inlet air flow rate of SOFC/MGT.

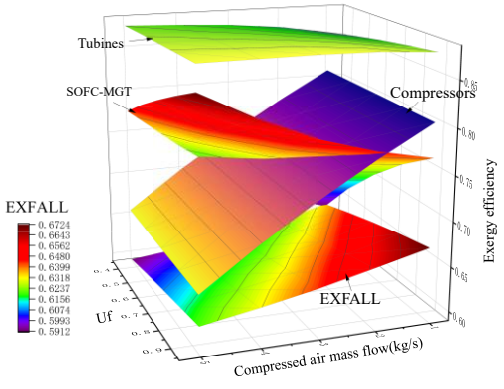


Figure 6. Effect of variation SOFC/MGT Uf and Inlet Air Mass Flow on TEE

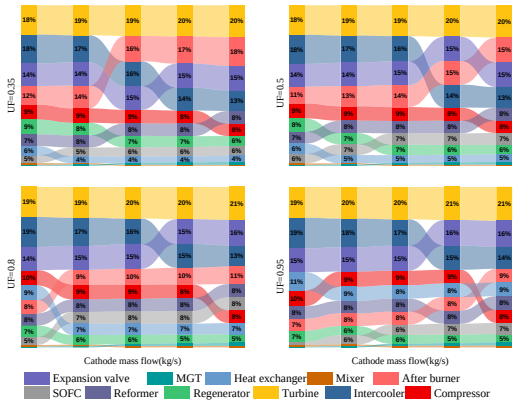


Figure 7. Percentage contribution of exergy destruction by each component with rise air mass flow to SOFC/MGT subsystem at varying fuel utilization factors.

4. Conclusions

This study presents a thermodynamic analysis of a novel hybrid CAES/SOFC/MGT generation system. A reliable static model of the system was established using ASPEN Customer Modeler and ASPEN Plus, and simulation experiments were conducted on key parameters to investigate the operating conditions of different subsystems and the overall system, in order to establish a reliable power generation and energy storage system. The following are some of the main results:

1. The influence of different parameters on the performance of hybrid system was evaluated by a parametric analysis. The study's findings suggest that to enhance the system's RTE and TEE, it is advisable to operate the system higher compressor and turbine isentropic efficiencies and optimal SOFC/MGT split air flow rates. These findings can be employed in future system design and modeling endeavors to perform multi-objective optimization, leading to superior system performance.

2. The effects of changes in multiple key system input parameters on efficiency and exergy efficiency were analyzed under different working modes and subsystems, which can better help to perform individual optimization of the actual operating conditions of the entire system.

The results offer valuable insights into the optimal operating conditions and design parameters for the system and can aid in the development of more efficient and cost-effective energy storage and generation solutions. Nevertheless, the static analysis approach proposed in this study provides valuable guidance for future system improvement and optimization. Furthermore, the model results obtained in this study not only serve as a valuable data source for dynamic model establishment, but also lay a foundation for multi-objective optimization of the system. Therefore, the value of system designed parameters should be selected wisely. Moreover, based on the findings of this study, the commercial potential of the system warrants further exploration.

Acknowledgment

This work was supported by the National Natural Science Foundation of China. (Grant number [51876011]).

Nomenclature

Abbreviations

CAES	Compressed air energy storage
SOFC	Solid oxide fuel cell
MGT	Micro gas turbine
AFC	Alkaline fuel cell
PAFC	Phosphoric acid fuel cells
MCFC	Molten carbonate fuel cell
PEMFC	Proton exchange membrane fuel cell
DMFC	Direct-methanol fuel cells
A-CAES	Adiabatic compressed air energy storage
AA-CAES	Advanced adiabatic compressed air energy storage
I-CAES	Isothermal compressed air energy storage
D-CAES	Diabatic compressed air energy storage

Parameters

P	Pressure (Pa)	Uf	Fuel utilization
T	Temperature(K)	E	Potential (V)
W	Work(J)	R	Molar gas constant ($J \cdot K^{-1} \cdot mol^{-1}$)
EX	Exergy (kW)	F	Faraday constant ($C \cdot mol^{-1}$)
h	enthalpy ($kJ \cdot kg^{-1}$)	i	Current density (A/cm^2)
π	Compression ratio	**Subscripts**	
k	Specific heat ratio	out	Outlet
m	Mass flow (kg/s)	in	Inlet
C_p	Heat capacity at constant pressure （J/K）	c	Compressor
Q	Amount of heat (kJ)	$inter$	Intercooler
λ	Efficiency	re	Regenerator
A	Area (m^2)	t	Turbine
μ	Thermal conductivity (W/m·K)	$Nernst$	Nernst Potential
V	Voltage (V)	act	Activation
η	Polarization loss(V)	con	Concentration
X	Mole fraction	ohm	Ohmic
α	Transmission coefficient	$split$	Split after 1st compressor
I	Exergy loss (kW)		

References

IPCC, (2022), Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change ,P.R. Shukla, J. Skea, R. Slade, A. Al Khourdajie, R. van Diemen, D. McCollum, M. Pathak, S. Some, P. Vyas, R. Fradera, M. Belkacemi, A. Hasija, G. Lisboa, S. Luz, J. Malley, (eds.). Cambridge University Press, USA. doi: 10.1017/9781009157926

Bazdar E, Sameti M, Nasiri F, et al. (2022) 'Compressed air energy storage in integrated energy systems: A review', *Renewable and Sustainable Energy Reviews*, pp. 167: 112701. doi: doi: 10.1016/j.rser.2022.112701.

Zhang, W., Xue, X., Liu, F., et al. (2020) 'Modelling and experimental validation f advanced adiabatic compressed air energy storage with off-design heat exchanger', *IET Renewable Power Gener*. 14, pp. 389–398. doi: 10.1049/iet-rpg.2019.0652.

G. Grazzini and A. Milazzo. (2008) 'Thermodynamic analysis of CAES/TES systems for renewable energy plants', *Renewable Energy*, vol. 33, pp.1998-2006. doi: 10.1016/j.renene.2007.12.003.

Javier, M., Ordóñezb, A., Álvarezb, R., et al. (2019) 'Energy from closed mines:underground energy storage and geothermal applications', *Renew. Sust.Energy Rev*, pp. 498–512. doi: 10.1016/j.rser.2019.04.007.

Li Y, Miao S, Yin B, Han J, Zhang S, Wang J, Luo X. (2019) 'Combined heat and power dispatch considering advanced adiabatic compressed air energy storage for wind power accommodation', *Energy Convers Manag*ement.pp 200: 112091. doi: 10.1016/j.enconman.2019.112091.

Nikolakakis, T., Fthenakis, V. (2017) 'The value of compressed-air energy storage or enhancing variable-renewable-energy integration: the case of Ireland',*Energy Technol*, pp. 2026–2038. doi: 10.1002/ente.201700151.

J. Nease and T.A. Adams II.(2014) '2014a, Coal-fuelled systems for peaking power with 100% CO2 capture through integration of solid oxide fuel cells with compressed air energy storage', J.*Power Sources*, pp. 92-107. doi: 10.1016/j.jpowsour.2013.11.040.

Nease, J., Monteiro, N., & Adams II, T. A. (2016). 'Application of a multiple time-scale rolling horizon optimization technique for improved load-following of an integrated SOFC/CAES plant with zero emissions', *In Computer Aided Chemical Engineering* , pp. 1725-1730. doi: 10.1016/b978-0-444-63428-3.50292-7.

Huang S, Yang C, Chen H, et al.(2022) 'Coupling impacts of SOFC operating temperature and fuel utilization on system net

efficiency in natural gas hybrid SOFC/GT system', *Case Studies in Thermal Engineering*, pp. 101-118. doi: 10.1016/j.csite.2022.101868.

Hartmann N, Vöhringer O, Kruck C, et al.(2012) 'Simulation and analysis of different adiabatic compressed air energy storage plant configurations', *Applied Energy*, pp. 541-548. doi: 10.1016/j.apenergy.2011.12.007.

Guo C, Xu Y, Zhang X, et al. (2017) 'Performance analysis of compressed air energy storage systems considering dynamic characteristics of compressed air storage', *Energy*, pp. 876-888. doi: 10.1016/j.energy.2017.06.145.

Guo H, Xu Y, Zhang Y, et al.(2019) 'Off-design performance and an optimal operation strategy for the multistage compression process in adiabatic compressed air energy storage systems', *Applied Thermal Engineering*, pp.262-274. doi: 10.1016/j.applthermaleng.2018.12.035.

Wu S, Zhou C, Doroodchi E, et al.(2019) 'Thermodynamic analysis of a novel hybrid thermochemical-compressed air energy storage system powered by wind, solar and/or off-peak electricity'. *Energy Conversion and Management*, pp.1268-1280. DOI: 10.1016/j.enconman.2018.11.063

Liese E A, Gemmen R S, Smith T P, et al.(2006) 'A dynamic bulk SOFC model used in a hybrid turbine controls test facility',*Turbo Expo: Power for Land, Sea, and Air*,pp.117-126. doi: 10.1115/gt2006-90383

Green production of dimethyl ether (DME) - indirect conversion of synthesis gas produced from biomass

Marianne S. Eikeland[a]*, Sebastian Larsen[a], Oliver Numme[a], Eivind J. R.Trasti[a], Terje Bråthen[a]

[a]*Department of Process, Energy and Environmental Technology, University of South-Eastern Norway (USN), Porsgrunn, Norway*
corresponding author: marianne.eikeland@usn.no

Abstract

In the transition to a fossil-free transport sector, the use of Dimethyl ether (DME) can be an environmentally friendly alternative. DME is a synthetically produced alternative to fuels like diesel or liquified petroleum gas (LPG), and has lower emissions of CO_2, NO_x and particles compared with diesel. To be a green renewable alternative, DME needs to be produced from carbon neutral resources such as biomass. DME can be produced from synthesis gas produced by gasification of biomass. The synthesis gas can be used to produce either DME directly in a single stage process with a bi-functional catalyst, or in a twostep process in which methanol is produced in the first step and converted to DME via dehydration in the second step. In this study process simulations of the DME synthesis from methanol is assessed. The paper involves assessment of process parameters and energy improvement of the DME synthesis. The study evaluates the effects of different thermodynamic models like PRSV, NRTL, WILSON and UNIQUAC in Aspen Hysys. Conversion reactor and Gibbs reactor configurations, and sensitive analysis of process parameters is studied. Heat integration is evaluated for energy resource management and cost estimation. The Gibbs reactor with the UNIQUAC model and internal heat integration resulted in an increase in DME production of 0.5% and a reduction in energy demand of 46%.

1. Introduction

The world's energy consumption is on the rise, and as greenhouse gas emissions continue to increase, a complete shift from fossil fuels to renewable energy sources has become imperative. The transport sector accounts for 15% of direct global greenhouse gas emissions (Energi og Klima 2023). The use of dimethyl ether (DME) can be an environmentally friendly alternative for the transition to a fossil-free transport sector, if DME is produced from carbon-neutral resources such as biomass. DME and diesel are fuels that can be used in the transport sector for the same purposes. DME offers some advantages over diesel, such as a higher cetane number which gives lower emissions, better cold start and lower consumption than a fuel with a lower cetane number. DME also produces lower emissions of particles, NOx, hydrocarbons, and carbon monoxide (CO) compared to diesel. However, the disadvantages of DME are lower density and viscosity which requires adaptations in the injection system (Salomonsson 2023). The volumetric energy density of DME and diesel are 21 MJ/dm^3 and 36,6 MJ/dm^3 respectively, indicating for every liter of diesel burned in an internal combustion engine there is a need of 1.74 liters of DME to achieve the same energy output.

Biomass can be used to produce energy-rich synthesis gas through a thermochemical process called gasification. The composition of the generated gas varies based on the type of biomass utilized, primarily comprising CO, H_2, CO_2, CH_4, and N_2. A pure synthesis gas predominantly comprises H_2 and CO, serving as essential building blocks for the production of biofuels or chemicals. Some examples of chemical products are ammonia, synthetic petrol and diesel, acetic acid, plastic raw materials, methanol and DME (Evans G 2012). The process of biomass gasification not only aids in waste reduction but also enables the productive utilization of materials like agricultural and forest residues.

The synthesis gas can be converted to DME in two ways; direct conversion (one-step process) or indirect conversion (two-step process), as shown in Figure 1. (U.S. Department of Energy 2023).

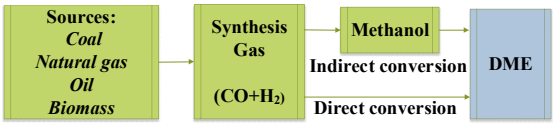


Figure 1: Dimethyl ether production diagram (Azizi Z. 2014).

The indirect conversion is a twostep process; first, methanol is made from synthesis gas in a reactor with a metallic catalyst. The methanol is dehydrated to DME in the next reactor with an acidic catalyst. In the direct conversion, DME is made in one step; the synthesis gas reacts directly to DME in a reactor with a dual catalyst system that combines the two reactions. This method has the advantage of avoiding the accumulation of methanol as an intermediate, thus increasing the yield of DME (Azizi Z. 2014). Both processes are mildly exothermic as shown in Table 1 (Guffanti S. 2021), thus it is important to control the temperature in the reactors to avoid the equilibrium reactions being reversed. (Polsen C. 2020) (Pagán-Torres Y. J. 2017). DME produced via methanol dehydration over acid catalysts such as γ-Al2O3, takes place at temperatures above 240 °C and pressures above 10 bar (Peinado C. 2020).
The reactions taking place in direct and indirect conversion is described in Table 1, where the last reaction is the main reaction taking place in the conversion of methanol to DME and water.

Table 1: Reactions and enthalpy change for the direct and indirect process (Guffanti S. 2021).

Reactions	**$\Delta H°298K$ [kJ/mol]**
$CO + 2H_2 \leftrightharpoons CH_3OH$	-90.4
$CO_2 + 3H_2 \leftrightharpoons CH_3OH + H_2O$	-49.4
$CO + H_2O \leftrightharpoons CO_2 + H_2$	-41.1
$2CH_3OH \leftrightharpoons (CH_3)_2O + H_2O$	-23.0

In this paper the focus is on the second step in the indirect conversion process which methanol is dehydrated to DME. It is assumed that methanol is available as a feed stream and that the first step, the methanol synthesis from syngas, has already been performed (Fossen M. 2022). Thus, only the reaction of methanol in a reactor to produce DME and water is simulated. Aspen Hysys simulations is used to study the process.
The objective of this research work is to assess process design, process parameter study and energy integration, as well as cost estimation of the DME synthesis.

2. Methodology

Figure 2 shows the process flow diagram for the DME synthesis simulated in Aspen Hysys (Turton R. 2018) (Larsen S. 2023).
Pure methanol (stream 1) is mixed with recycled reactant stream (stream 13) and evaporated in the heater E-201 before being fed to the reactor R-201. The reactor is set to operate at a temperature of 250 °C and 14,7 bar. The effluent from the reactor (stream 6) is cooled by the cooler E-203 before entering the first distillation column T-201 at 89°C and 10,4 bar. In this column the DME is the distillate and final product, and methanol and water are the bottom product. The methanol mixture is then cooled down in the cooler E-205, before it enters the second distillation column T-202, where the water is separated from methanol as the bottom product and is sent to wastewater treatment to remove traces of organics. Unreacted methanol in the distillate is recycled and combined with pure methanol in the mixer M-201. The purity of the DME in the first distillation column is set to 99.5%. The component recovery of methanol was set to 95% in the second distillation column. Simulated with the thermodynamic model PRSV (Peng-Robinson-Stryjek-Vera), this model is referred to as the base case.

2.1 Process parameter study

The main reaction of methanol to DME and water is an exothermic reaction. Increasing the temperature for this reaction the system will consume some of the heat by shifting the equilibrium to the left as described by the Le Chatelier's principle. This will lead to a reduction of the DME concentrations when the temperature increases. Since the main reaction has the same number of molecules on both sides of the equilibrium, the pressure has no or little effect on the composition of DME.

2.2. Thermodynamic models

Four thermodynamic models were studied in this work, all of them relevant for the synthesis of DME. These are PRSV, UNIQUAC, NRTL and Wilson. The thermodynamic models calculate physical and transport properties, as well as phase behavior for the simulated processes. These are the models recommended from Aspen Hysys for these types of polar mixtures. AspenTech (Aspen Technology, Inc 2020) describes the thermodynamic models as follow:
PRSV (Peng-Robinson-Stryjek-Vera) considers moderate deviations from ideality in gas and non-ideal liquid phases. This model is well suited for aqueous solutions containing water, methanol, or glycols.
UNIQUAC (Universal Quasi Chemical) is a model that describes the liquid structure using activity.

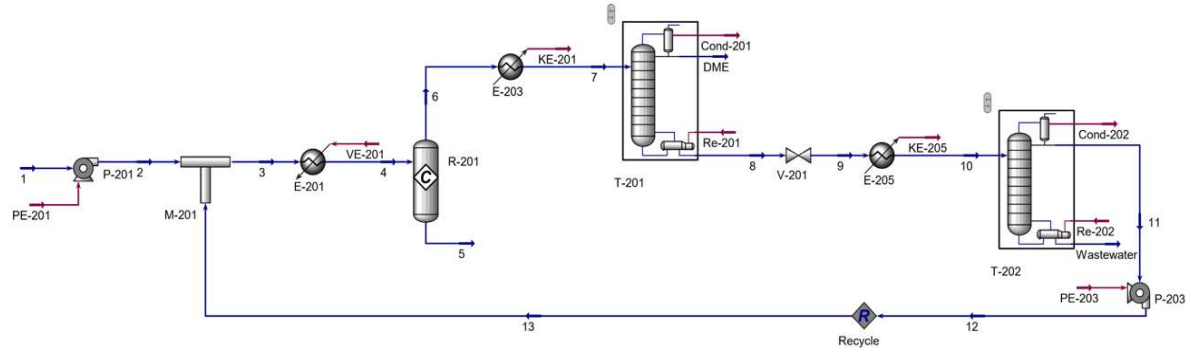


Figure 2: Process flow diagram for the base case of the synthesis of dimethyl ether (DME) in Aspen Hysys (Larsen S. 2023).

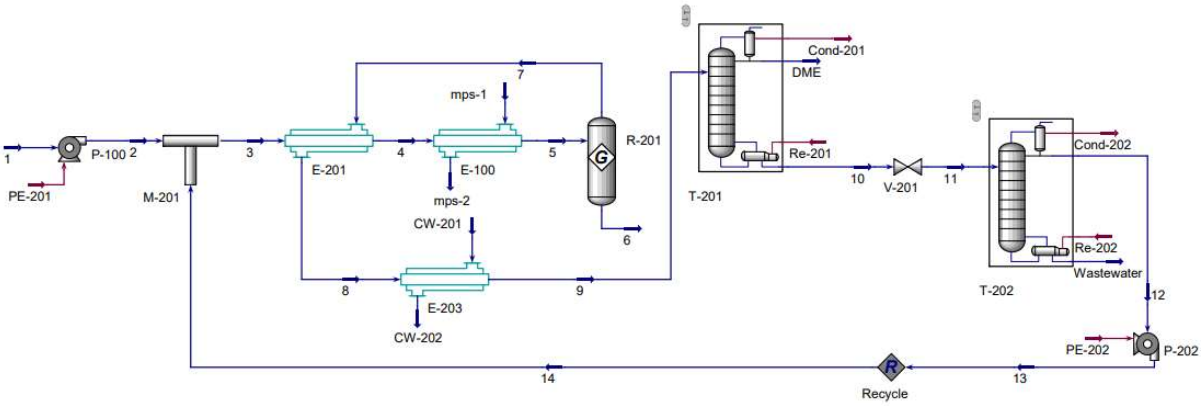


Figure 3: Process flow diagram with improved energy integration of the synthesis of dimethyl ether (DME) in Aspen Hysys (Larsen S. 2023).

These activity coefficients are factors that can account for deviations from ideality in the liquid phase at low to moderate pressures.
NRTL (Non-Random-Two-Liquid) also uses activity coefficients to represent non-ideality in the liquid phase. It can also handle electrolytes and gases with pressures above 10 bar. NRTL can calculate properties for pure components such as methanol, water and DME.
Wilson is also using activation coefficients, but with a simpler form than UNIQUAC and NRTL. It cannot be used for systems with two liquid phases.

2.2 Reactor design

Equilibrium reactions in Aspen Hysys should be performed in conversion reactors, or in Gibbs reactors. A conversion reactor is a simple type of reactor in Aspen Hysys where the conversion proportion is specified. This reactor needs a reaction kit that contains conversion reactions and operates on a stoichiometric basis. The reaction stops when the specified conversion is achieved. The advantage of this reactor is that it can integrate several different reactions in the same unit. This can be useful in extremely complex reactor designs (Hafiza Shukor P. 2023).
The reactions in a Gibbs reactor determines the effluent composition by achieving phase and chemical equilibrium. This reactor does not need to use reaction stoichiometry but calculates the product mix by minimizing the Gibbs free energy of the system. In the Gibbs reactor, we assume that neither pure substances nor the reaction mixture is ideal. An advantage of this reactor is that it can act as a separator, an equilibrium reactor, or a reactor that minimizes Gibbs free energy without specifying any reactions. The Gibbs reactor can be useful for simulating the equilibrium state or when reaction kinetics are lacking. It is also easier to converge than an equilibrium reactor with reactions when the simulation is sensitive to the input parameters (Hafiza Shukor P. 2023).

2.3 Energy integration

For energy improvement, the thermodynamic model that gave the highest DME mass production was used. However, in practical scenarios, it is crucial to validate the choice of a thermodynamic model by comparing it with experimental data. This validation process ensures that the selected model accurately represents the real-world situation.
A heat exchanger E-302, given in Figure 3, replaced the heater E-201 in Figure 2. The reaction in the reactor R-301 was exothermic, heat was generated and released. This means that the outlet flow from the reactor was hotter than the inlet flow. This heat

flow was used on the shell side of the heat exchanger E-301 in Figure 3 so the heat exchanger E-302 had a reduced energy requirement. The cooler E-203 was replaced with a heat exchanger E-303 which used cooling water to lower the temperature further before the stream enters the distillation column T-301. The second cooler E-205 was removed completely without replacement, because the temperature had been lowered by changing the operating parameters earlier, thus no need of cooling was necessary. The changes for energy improvement and operating parameters are referred to as the energy integrated case.

2.4 Cost estimation

Cost estimation was done for both investment cost (CAPEX) and operating cost (OPEX).

Calculation of investment costs was performed using Equation 1, where the CAPEX represents the investment costs. ΣC_{BM} is the sum of the purchase cost and installation cost for all appliances and equipment in the DME synthesis.

$$CAPEX = 1{,}18 * \Sigma C_{BM} \quad (1)$$

OPEX consists of three categories: direct production costs (DMC), fixed production costs (FMC), and general production costs (GE). To calculate the electricity cost, an average price of 0.192 USD/kWh for global businesses in 2022 was used (GlobalPetrolPrices.com 2022). For calculating the cooling water cost, a price of 0.0157 USD/m^3 in the year 2001 (Turton R. 2018) was used. To calculate the steam cost, prices from (Turton R. 2018) were used, with a medium steam pressure price of 0.00961 USD/m^3 and a high steam pressure price of 0.00954 USD/m^3. To calculate the raw material cost, a price of 1.49 USD/US gallon for methanol was used (Seaberg 2018).

The calculation of operating expenses (OPEX and COM_d) excluding depreciation was performed using Equation 2, where FCI represents the investment cost, C_{OL} denotes operating labor cost, C_{UT} signifies the cost of utilities, C_{WT} represents the cost of wastewater management, and C_{RM} represents the cost of raw materials.

$$OPEX = COM_d = 0.180*FCI + 2.73*C_{OL} + 1.23*(C_{UT} + C_{WT} + C_{RM}) \quad (2)$$

For the calculation of DME income (R), an average price (P) for DME in 2018 of 2.26 USD/US gallon was used (Seaberg 2018). This price is competitive with diesel considering the ratio between them of 1.74. The formulas and data for the cost estimation of CAPEX and OPEX are given in (Turton R. 2018). The cost was estimated for the base case, for the process parameter study, and the energy integrated process.

To calculate the net present value (NPV), a cumulative discounted cash flow was computed throughout the construction and operational phases of the DME synthesis. NPV is a measure of the project's profitability based on the present value of all incomes and expenses associated with the project. The cumulative discounted cash flow is the sum of all future cash flows (incomes - expenses) generated and accumulated, adjusted back to year zero.

3. Results

3.1 Process parameter study

In Figure 4, the production of DME was studied as a function of temperature and pressure. The mole fraction of DME in the outlet flow of the reactor as a function of temperature was studied in the temperature range 200-370°C at 14.7 bar pressure. While the mole fraction as a function of pressure were studied at 250°C in the pressure range 5-30 bar. The simulations were performed with the thermodynamic model PRSV.

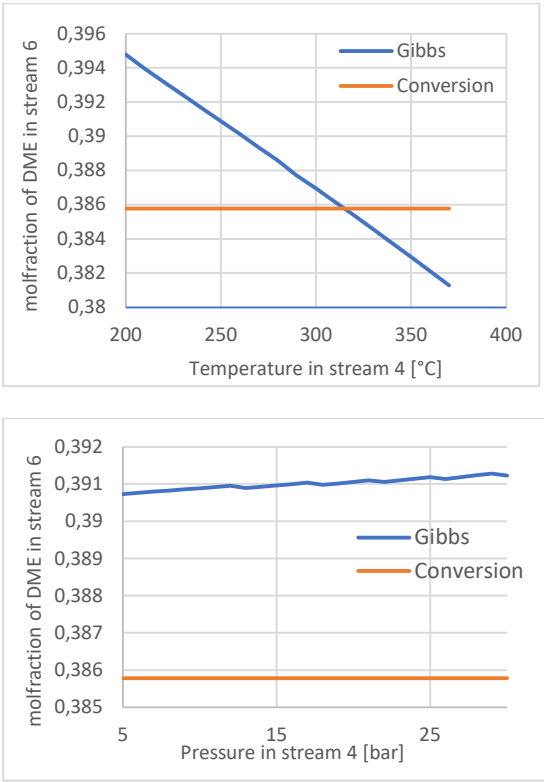


Figure 4: Mole fraction in stream 6, outlet flow of the reactor, as function of temperature and pressure, for the conversion and Gibbs reactor design.

The reaction efficiency in the conversion reactor was defined to be constant to 80%. Changing the pressure and temperature in stream 4, the inlet stream to the reactor, did not lead to a change in the DME stream mole fraction, which is reasonable as the conversion efficiency calculation was both

temperature and pressure independent. Therefore, this reactor was not assessed further.
For the Gibbs reactor, the DME mole fraction in the outlet decreases with higher temperature in the inlet. The chemical reaction is exothermic and will occur spontaneously. This results in a negative value of the Gibbs free energy and a negative enthalpy change. The reaction will be reserved to stop the temperature increase resulting in a lower DME production. An increase in pressure favors the production rate, because the first 2 reactions in Table 1 contributes to larger methanol production. However, the changes in the tested pressure range are minimal.
The Gibbs reactor at 200°C and 17 bar was tested to confirm the impact of lower temperature and increased pressure on energy improvement and increased DME production. Table 2 shows the changes in energy consumption as a result of parameter changes over the reactor system. A decrease in energy consumption of 6.6 % is observed.
A parameter evaluation was also performed over the first distillation column T-201. The temperature inlet of flow 7 was increased from 89°C to 135°C, the pressure remains as in base case at 10.4 bar.

Table 2: Simulation results for the Gibbs-reactor parameter adjustment with PRSV.

	Base case		**Parameter change**	
Flow	4	6	4	6
Temperature [°C]	250	363,9	200	318,3
Pressure [bar]	14.7	14.7	17	17
Mass flow DME [kg/h]	0	5 905	0	5 913
Energy consumption in the system[kW]	17 652		16 491	
Change in mass flow [%]			0.13	
Change in energy [%]			6.6	

Table 3: Simulation results for the parameter adjustment in the distillation column T201 with PRSV.

	Base case		**Parameter change**	
Flow	7	DME	7	DME
Temperature [°C]	89	45,78	135	45,78
Pressure [bar]	10.4	10.3	10	10.3
Mass flow DME [kg/h]	5 913	5 913	5 913	5 914
Energy consumption in the system[kW]	16 491		14 098	
Change in mass flow [%]	0		0.02	
Change in energy [%]			14.5	

Table 3 give the result of the simulation over the column. It is observed a reduction of energy of 14% in the system using an increased temperature into the column because of reduction in cooling water. A total energy reduction of 20% are observed because of parameter adjustments.

3.2. Thermodynamic models

Based on Gibbs reactor, different thermodynamic models was simulated for comparison. Figure 5 shows an overview of the change in energy and the outlet flow of DME for the four thermodynamic models compared to the base case. The UNIQUAC model gave the lowest energy consumption before energy integration. For the mass flow an increase of 0.5% for the UNIQUAC model is observed compared to the base case.
The UNIQUAC model is well suited for gases with high pressure and a known boiling point. The UNIQUAC model was also recommended by specialist literature (Turton R. 2018). In a real process, it is essential to choose the thermodynamic model that best describes the given process. This could be done by comparing experimental data or real industry data with simulation results. For this work, process data was not available to carry out validation of the simulation model.

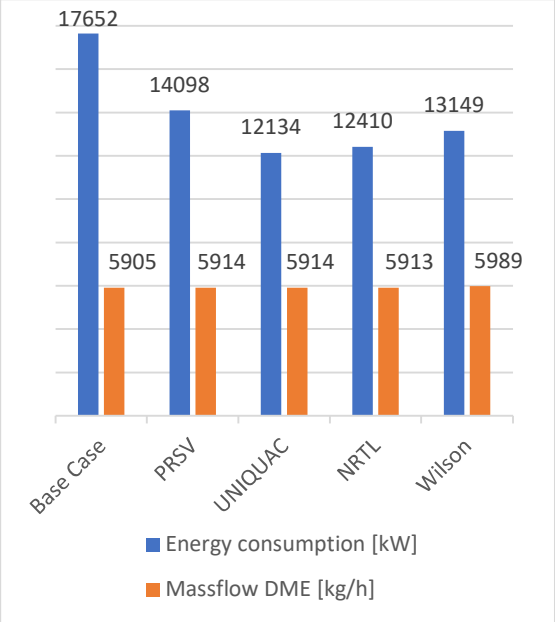


Figure 5: Comparison of energy consumption[kW] and mass flow DME [kg/h] for different thermodynamic models.

3.3 Energy integration

The UNIQUAC model has been chosen when simulating the energy integrated process. Figure 6 shows an overview and comparison between the base case, the process parameter study, and the energy integrated process. For the process parameter study the pressure and temperature are changed to 17 bar and 200°C in the reactor, and UNIQUAC is used as the thermodynamic model. For the energy integrated process two heat exchangers are inserted,

one was inserted with internal heat integration to reduce energy consumption in the heater. The second heat exchanger used external heat transfer in the form of cooling water as a replacement for two coolers. An energy reduction of 46% is observed from the base case to the energy integrated case.

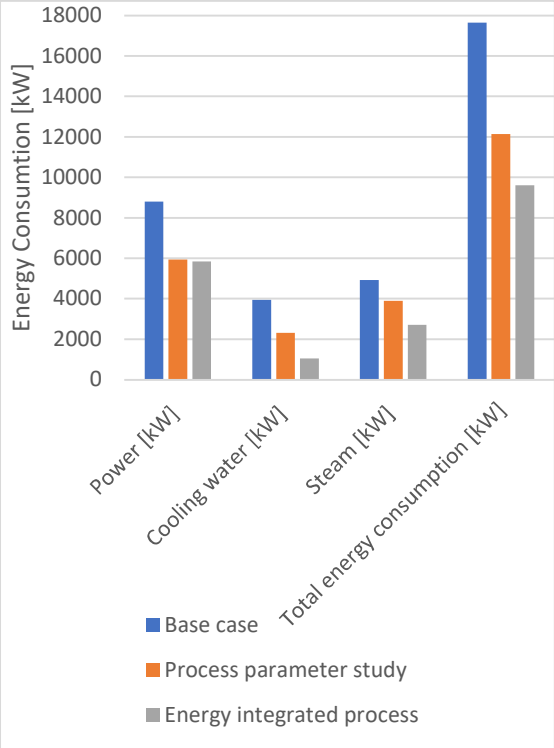


Figure 6: Comparison of energy consumption between base case, process parameter study and energy integrated process.

3.4 Cost estimation

By reducing the energy requirement, the OPEX for the energy was reduced. However, as pressure and temperature were not changed much compared to the base case, CAPEX for the equipment was approximately the same. The total OPEX was reduced with -11%, while the CAPEX was reduced with -4% compared to the base case. The project estimated the CAPEX to approximately $1.8 million, and the OPEX of approximately $58.6 million, and a revenue of approximately $54.3 million.

For the utilities, the costs of electrical energy, cooling water and steam were reduced from the base case, as shown in Table 4, by 34%, 74% and 49% respectively. This is due to lower temperature and higher pressure, which led to a lower need for energy in pumps, evaporators, and condensers altogether. It is also due to the internal heat integration that utilized the heat from the reactor to heat the inlet stream. This led to less need for steam, which also required lower pressure and thus a lower price. An increased temperature before the first distillation column T-301, led to less need for cooling and removal of the second cooler E-205. This resulted in less need for cooling water. Total savings for utilities were 35% compared to the base case.

The pressure of the reactor was earlier set to 17 bar to examine the impact of the pressure on DME production and cost estimation. Because of small impact on DME production and cost savings, the pressure in the reactor was reduced to 10 bar. This pressure reduction reduced the cash flow with only 0.02 % or approximately 11 000 USD/year. This shows that reduction in the reactor feed pressure has insignificant impact on the cost estimation, but it is overall beneficial for safety issues.

Table 4: Overview of savings in utility cost per year and percentage for changes from the base case to the energy-integrated process.

	Change [USD/year]	**Change [%]**
Electrical energy	-5 000 000	-34
Cooling water	-48 000	-74
Steam	-850 000	-49
Totalt	-5 900 000	-35

During the evaluation of the net present value (NPV) reaching zero in year 25 at a discount rate of 10%, it was determined that the DME price needed to be 2.46 USD/US gallon in order to achieve the breakeven point.

4. Summary and Discussions

This paper used Aspen Hysys simulations to describe the production of green DME from indirect conversion of synthesis gas from biomass. The main aim was to evaluate process design, process parameter study and energy improvement, as well as cost estimation of the DME synthesis. The aim was to find a process with improved DME yield and energy requirement with minimal adjustments to operating parameters.

The process has been simulated and improved using various operating parameters, reactor design and thermodynamic models. The final process configuration was found to be a Gibbs reactor with the UNIQUAC model and internal heat integration. This process resulted in an increase in the DME production of 0.5% and a reduction in energy demand of 46%. The cost estimate showed that the project was not profitable with the given prices for methanol and DME of 1.49 and 2.26 USD/US gallon respectively. The price for DME had to be at least 2.46 USD/US gallon to get a positive NPV within 25 years. By setting the price at 2.50 USD/US gallon the repayment period will be approx. 6.5 years. The price of DME (Dimethyl Ether) and Methanol is indeed crucial for profitability. The choice of a discount rate of 10% also significantly impacts the

result, as operating expenses (OPEX) are much larger compared to capital expenses (CAPEX). Additionally, utility costs are also significant, and there is potential for improvement in terms of heat integration. The price of electricity, fixed at 0.19USD/kWh, will also have a considerable impact on the result.

Based on the findings, it was concluded that production of green DME from synthesis gas is technically and economically feasible.

It is also concluded that DME is an attractive fuel for the heavy transport sector and municipalities. Further research into parameter changes and energy integration is necessary to optimize the process profitability.

References

Aspen Technology, Inc. 2020. «Aspen Hysys V 12, Aspen HYSYS Help, Fluid Package Referances.»

Azizi Z., Rezaeimanesh M., Tohidian T., Rahimpour M. R. 2014. «Dimethyl ether: A review of technologies and production challenges.» *Chemical Eengineering and Processing: Process Intensification ,Volum 82*, 150-172. doi:10.1016/j.cep.2014.06.007.

Basu, Prabir. 2013. *Biomass Gasification, Pyrolysis and Torrification, Practical Design and Theory.* Second edition. Academic press.

Energi og Klima. 2023. “Utslipp sektorer.” *https://energiogklima.no/klimavakten/utslipp-sektorer/.* 29 Mar.

Evans G, Smith C. 2012. “Biomass to Liquids Technology.” In *Comprehensive Renewable Energy*, by Ali Sayigh, 155-204. Elsevier Ltd. doi:10.1016/B978-0-08-087872-0.00515-1.

Fossen M., Halvorsrød J., Narvestad T., Tjemsland S. , Timsina R., Eikeland M. 2022. «Aspen Hysys simulation of the methanol synthesis based on gas from biomass gasification.» *Conference: 63rd International Conference of Scandinavian Simulation Society, SIMS 2022, Trondheim, Norway, September 20-21, 2022.* Linköping Electronic Conference Proceedings 192. doi:10.3384/ecp192052.

GlobalPetrolPrices.com. 2022. *GlobalPetrolPrices.com.* September. Accessed June 2, 2023. https://www.globalpetrolprices.com/electricity_prices/.

Guffanti S., Visconti C. G., Groppi G. 2021. «Model Analysis of the Role of Kinetics, Adsorption Capacity, and Heat and Mass Transfer Effects in Sorption Enhanced Dimethyl Ether Synthesis.» *Ind. Eng. Chem. Res., vol. 60, no. 18, pp. 6767-6783.* doi:10.1021/acs.iecr.1c00521.

Hafiza Shukor P., Syahidah Ku Ismail D. K., Hafizah Mohad Jonar C. 2023. «ERT 214 MATERIAL AND ENERGY BALANCE HYSYS SIMULATION MANUAL.» 12 Apr. https://docplayer.net/40110495-Ert-214-material-and-energy-balance-hysys-simulation-manual.html.

Larsen S., Numme O., Trasti E. J. R. 2023. *Grønn produksjon av dimetyleter (DME) -indirekte konvertering av syntesegass fra biomasse.* Bachelor Thesis, Porsgrunn: University of South-Eastern Norway.

Pagán-Torres Y. J., Lu J., Nikolla E., Alba-Rubio A. C., Farrusseng D., Tuel A. 2017. “Well-Defined Nanostructures for Catalysis by Atomic Layer Deposition.” In *Studies in Surface Science and Catalysis*, by Cargnello M Fornasiero P., 643-676. Elsevier B.V.

Peinado C., Liuzzi D., Ladera-Gallardo R.M., Retuerto M., Ojeda M., Peña M. A., and Rojas S. 2020. «Effects of support and reaction pressure for the synthesis of dimethyl ether over heteropolyacid catalysts.» *Scientific Reports, Natureresearch.* doi:https://doi.org/10.1038%2Fs41598-020-65296-3.

Polsen C., Narataruksa P., Hunpinyo P., and Prapainainar C. 2020. «Simulation of single-step dimethyl ether synthesis from syngas.» *Energy Rep., vol. 6.* doi:10.1016/j.egyr.2019.11.112.

Salomonsson, Per. 2023. «BioDME , About DME.» *http://www.biodme.eu/about-dme.* 26 May.

Seaberg, Josh. 2018. *Dimethyl ether for transportation.* Honors College Theses, AIChE Design. https://shareok.org/handle/11244/302127.

Turton R., Shaeiwitz J. A., Bhattacharyya D., Whitting W. B. 2018. *Analysis, synthesis, and design of chemical process.* Pearson Education Inc.

U.S. Department of Energy . 2023. “Energy Efficiency & Renewable Energy, AFDC Alternative fuels Data Center: Dimethylether.” *https://afdc.energy.gov/fuels/emerging_dme.html.* Feb.

Enhancing Indoor Environmental Simulations: A Comprehensive Review of CFD Methods

Shahrzad Marashian[a,*], Amir Vadiee [a], Omid Abouali [b], Sasan Sadrizadeh [a,b,*]

[a] *School of Business, Society, and Engineering, Mälardalens University, Västerås, Sweden,*

[b] *Department of Civil and Architectural Engineering, KTH Royal Institute of Technology, Stockholm, Sweden*
shahrzad.marashian@mdu.se | sasan.sadrizadeh@mdu.se

Abstract

Computational Fluid Dynamics (CFD) simulations are extensively used to model indoor environments, including airflow patterns, temperature distribution, and contaminant dispersion. These simulations provide valuable insights for improving indoor air quality, enhancing thermal comfort, optimizing energy efficiency, and informing design decisions. The recent global pandemic has emphasized the importance of understanding airflow patterns and particle dispersion in indoor spaces, highlighting the potential of CFD simulations to guide strategies for improving indoor air quality and public health. Consequently, there has been a significant increase in research focused on studying the transport and dispersion of pollutants in indoor environments using CFD techniques. These simulations are vital in advancing engineers' understanding of indoor environments; however, achieving accurate results requires careful method selection and proper implementation of each step. This paper aims to review the state-of-the-art CFD simulations of indoor environments, specifically focusing on strategies employed for three main simulation components: geometry and grid generation, ventilation strategies, and turbulence model selection. Researchers can select suitable techniques for their specific applications by comparing different indoor airflow simulation strategies.

1. Introduction

Understanding and assessing indoor environments is crucial, given that people spend most of their time indoors. The quality of indoor air, thermal comfort, and energy consumption are key factors that directly impact individuals' well-being and productivity. Poor air quality can lead to various health issues, such as respiratory problems and allergies. After the recent pandemic, there has been a heightened awareness and understanding of the importance of indoor air quality. A growing emphasis is on implementing strategies and technologies that enhance ventilation, filtration, and air purification to create safer and healthier indoor environments. The COVID-19 pandemic has significantly increased research on indoor airflow since the onset of the COVID-19 pandemic in 2019 (Kohanski et al., 2020).

Investigating enclosed environments commonly involves conducting experiments and utilizing computational simulations. The increasing number of numerical research studies in this area can also be attributed to a notable advancement in computer capabilities, as well as the development and refinement of computational fluid dynamics (CFD) methods and software.

Indoor environments can be broadly categorized into private residential spaces and public settings. Public utility buildings include diverse spaces, such as churches, museums, libraries, and hospitals. Each of these environments serves a unique purpose and has specific requirements for indoor air quality. The quality of indoor air in public utility buildings is influenced by various environmental factors, including human activity, characteristics of the indoor area, and the presence of chemical compounds in the surrounding air (Śmiełowska et al., 2017).

Ventilation is an important part of indoor air simulations, which also plays a crucial role in limiting the spread of viruses. Optimizing ventilation rates, eliminating air recirculation, using portable air cleaners with proper maintenance, and avoiding overcrowding in public spaces are some of the recommendations in this area. Implementing these engineering controls alongside other preventive measures will lower airborne pathogen concentrations and reduce infection rates for airborne diseases. It emphasizes the need to prioritize airborne transmission reduction in hospitals and public buildings to protect healthcare workers and the public.(Morawska et al., 2020)

The effectiveness of designing and operating indoor environments relies on accurate and reliable numerical simulations. This paper comprehensively reviews the latest advancements in Computational Fluid Dynamics (CFD) simulations for various indoor settings. Specifically, it focuses on three crucial and challenging aspects: geometry and grid generation, ventilation strategy, and selection of appropriate turbulence models. By examining these areas, this review aims to enhance the understanding of simulation strategies and ultimately improve the overall accuracy and reliability of numerical simulations in the investigation of indoor environments.

2. Methodology

The method used in this study involves gathering information from recent articles by searching for keywords such as indoor airflow, ventilation, CFD simulation, and indoor air quality in academic databases such as Web of Science, SAGE journals, and Science Direct. The focus is on studies conducted since 2012. The collected information is then compared and analyzed, specifically emphasizing geometry and grid generation, ventilation strategies, and the selection of turbulence models. By examining and summarizing these aspects, this study aims to understand the latest advancements and strategies used in simulating indoor environments.

3. Overview of CFD Methods for Indoor Environment Simulations

3.1. Yearly publication distribution

For this review, an analysis was performed on 25 previous research studies in the field of CFD simulation with a specific emphasis on indoor airflow. *Figure 1* illustrates the yearly distribution of reviewed papers. It highlights the observed trend in the selected research, showing a focus on recent articles from 2020 until the present.

3.2. Simulation tools

The software tool ANSYS Fluent is widely recognized and extensively used for conducting CFD simulations in most research papers reviewed here. Fluent has gained popularity among researchers and engineers for its exceptional capabilities and user-friendly interface, making it a preferred choice for modeling and analyzing fluid flow, heat transfer, and other related simulations. This software is also commonly used for simulating indoor airflows and conducting CFD analyses in indoor air quality and ventilation. It gives researchers and engineers the tools to model and analyze various aspects of indoor airflows, including air distribution, temperature profiles, pollutant dispersion, and ventilation effectiveness. The choice of CFD software depends on the specific research objectives and the researchers' expertise. A few studies also employed other CFD tools for their simulations. For instance, some researchers utilized ANSYS CFX (Kalliomaki et al., n.d.), OpenFOAM (Arpino et al., 2023), or STAR-CCM+ (Chang et al., 2023).

3.3. Geometry and grid generation

Figure 2 categorizes the selected papers based on the different types of indoor spaces studied using CFD. These categories include various indoor environments, including Hospital Environments, Transportation Spaces, Educational Spaces, Offices, Restaurants, Residential Spaces, and Museums. Among these categories, Figure 2 shows that hospital ward research represents the most significant proportion, accounting for 36% of all the selected research studies. *Table 1* provides detailed information on the subsections within each category and lists the corresponding relevant research studies.

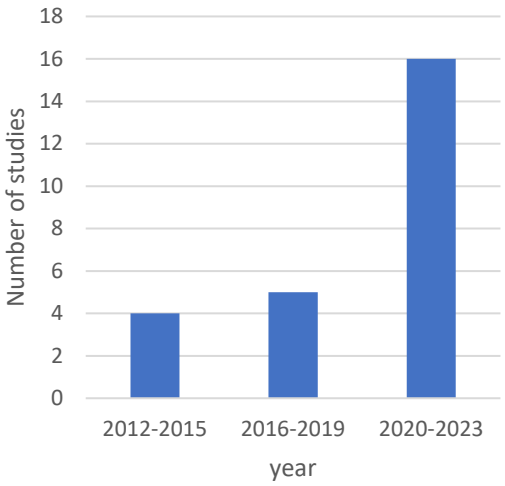


Figure 1: Yearly distribution of reviewed studies

Grid generation is a vital aspect of conducting indoor airflow CFD simulations. It involves creating a grid or mesh that discretizes the indoor environment into smaller elements, allowing for the numerical solution of governing equations. (Liu et al., 2022a) conducted a comparison of three different meshing strategies, namely hexahedral, tetrahedral, and polyhedral meshes, for simulating indoor airflow in geometries with varying levels of complexity to assess the performance and suitability of each mesh type in capturing the airflow behavior within indoor environments

All the articles studied here utilized three-dimensional geometry, and most employed unstructured grids due to their complex geometry. Various types of cells are used in grid generation in studied articles, including Tetrahedral, Hexahedral, poly-hexcore.

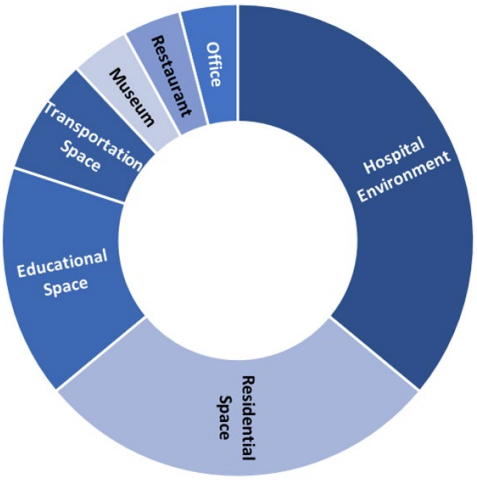


Figure 2: Different Types of Indoor Spaces in Reviewed Studies and Their Proportions

As depicted in *Figure 3*, half of the studies utilized a tetrahedral mesh for their simulations. Around 20% of the studies used hexahedral mesh. Only 10% of recent studies used a poly-hexcore mesh, which is also a new option available in recent versions of Fluent meshing software.

The volume of the simulated geometry varied across the different studies. The smallest volume, 0.128 m^3, was associated with a modeled room (Marashian et al., 2022), while the largest volume, 724.56 m^3, was observed in the context of an open museum space (Bakry et al., 2022).

The average mesh density is another grid property determined based on the ratio of the geometric model's volume to the total number of cells. (Liu et al., 2022a). This parameter also can represent the complexity of the generated grid. In Table 1, the average mesh density is calculated and reported for the studied papers.

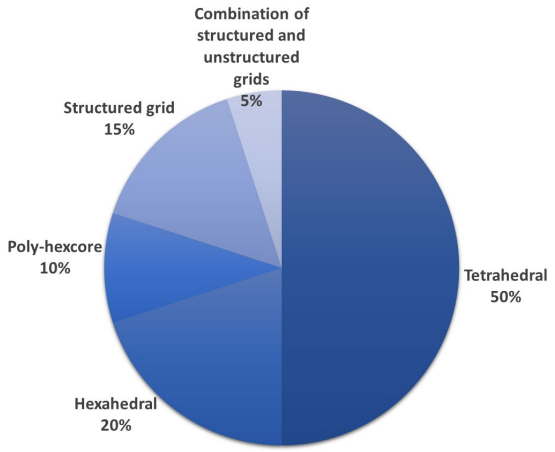


Figure 3: Cell Types Used in Grid Generation in Studied Articles

3.4. Ventilation strategies

Natural and mechanical ventilation are the two main methods to ventilate indoor spaces. Natural ventilation relies on natural forces such as wind or buoyancy to create airflow. On the other hand, mechanical ventilation utilizes mechanical systems like fans, blowers, or air conditioners to control indoor airflow.

The number of inlets, outlets, and openings like windows and their locations are crucial factors that significantly influence indoor air simulations. These parameters are critical in determining air distribution, ventilation effectiveness, and contaminant removal within indoor environments. The placement of inlets and outlets affects the airflow patterns and the distribution of fresh air within the indoor environment. Strategic placement of supply air diffusers and exhausts can ensure efficient air mixing, reduce the residence time of contaminants, and promote thermal comfort.

According to (Xu et al., 2022), Mixing ventilation (MV) and Displacement ventilation (DV) are two common approaches for distributing air within indoor environments. In mixing ventilation, high-speed air is released from upper diffusers, causing the supplied air to mix with the surrounding air. Displacement ventilation, on the other hand, involves supplying cool air from lower diffusers, utilizing convective thermal flow around heat sources, and expelling it from the top of the room. Personalized ventilation (PV) is another ventilation strategy that has recently been used for indoor environments. In personal ventilation, fresh air can be directly supplied for inhalation purposes or exhaled aerosols can be directly exhausted from their source. Downward ventilation (DWV), Protected zone ventilation (PZV), and Stratum ventilation (SV) are among other strategies used in literature.

Air Changes per Hour (ACH) is an important term of ventilation design that refers to the number of times the entire volume of air within a space is exchanged with fresh air per hour. ACH is a key parameter for quantifying ventilation rates and determining indoor air quality. Increasing the ventilation rate alone does not always ensure better contamination control. Therefore, it's important to consider the ventilation airflow pattern and the efficiency of air changes in order to achieve effective contamination control (Wang et al., 2018a).

Table 1 provides comprehensive details regarding the type of room, the ventilation strategy employed, and the Air Changes per Hour (ACH) values, wherever available, for the studies examined in this review. ACH values in the reviewed studies varied significantly, ranging from 0.5 for one of the residential building cases to as high as 100 for specific cases, such as operating rooms.

3.5. Turbulence models

Turbulence modeling is crucial in simulating indoor airflow in CFD simulations. Indoor environments are often characterized by complex flow patterns, including turbulence, which can significantly impact factors such as air quality, thermal comfort, and energy efficiency.

Researchers and engineers can better understand and optimize indoor airflow conditions by accurately modeling turbulence. The main turbulent models commonly used in CFD simulations are Direct Numerical Simulation (DNS), Large Eddy Simulation (LES), and Reynolds-Averaged Navier-Stokes (RANS). DNS resolves all the relevant spatial and time scales of the flow and none of the eddies are modeled but it needs a very high computational cost and resources. Hence it is mainly used for basic flows in simple geometries.

RANS models, such as the k-ε, k-ω, and their variations, are widely used in indoor airflow simulations due to their computational efficiency compared to DNS and LES. RANS models solve the time-averaged Navier-Stokes equations and provide insights into the mean flow characteristics.

The renormalization group k-ε turbulence model (RNG), realizable k-ε turbulence model, standard k-ε turbulence model, and Shear stress transport model k-ω(SST) are among the most employed RANS turbulence models in indoor airflow simulations. These models involve the estimation of turbulence kinetic energy (k) and its dissipation rate (ε or ω) to calculate the turbulent viscosity and model the turbulence effects.

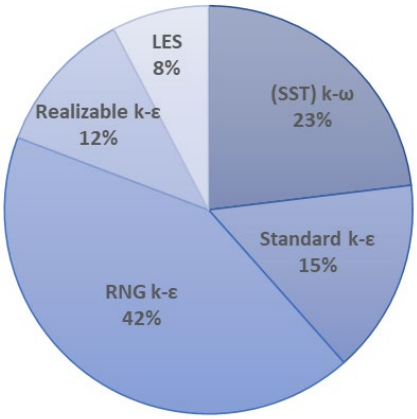


Figure 4: Turbulent models Used in selected Articles.

According to *Figure 4*, (RNG) k-ε was the most prevalent turbulence model among the reviewed studies, accounting for 42% of the cases. LES is a turbulence modeling approach that resolves the larger eddies in the flow while modeling the smaller dissipative scales. Due to its high computational expense, this model was used less frequently compared to other turbulence models and represents 8% of the models used in the reviewed studies.

Table 1: Geometry Category and Relevant Research Studies; Type of Room; Ventilation Strategy, and ACH (Where Available); Average Mesh Density; Turbulent Model

Types of Indoor Spaces		Study	Ventilation Strategy	ACH	Average Mesh Density cell/m3	Turbulent model
Hospital Environments	Patient ward	(Almhafdy et al., 2023)			29167	SST k-ω
		(Satheesan et al., 2020)	Mechanically ventilated (with a positive pressure towards the corridor), with and without local exhaust grilles	3 6 9 13		RNG k-ε
		(Lu et al., 2020)	Stratum ventilation Mixing ventilation Downward ventilation Displacement ventilation	12	8283	RNG k-ε
		(Aganovic et al., 2019)	Protected occupied zone ventilation	1.57 2.36 3.15 3.94 4.73	108696	SST k-ω

	Operating room	(Sadeghian et al., 2022)			40000	RNG k-ε / Realizable k-ε
		(Rahman et al., 2018)			6667	RNG k-ε
		(Wang et al., 2018)	Vertical laminar airflow ventilation	26	28925	Realizable k-ε
			Temperature-controlled airflow	46		
			Mixing ventilation	100		
		(Sadrizadeh et al., 2014)	Mixing ventilation system	47	19048	RNG k-ε
	Isolation-room	(Kalliomaki et al., 2020.)	Overhead mixing ventilation	12	46939	SST k–ω LES
			Local downward ventilation with background mixing ventilation			
			Zonal downward ventilation			
	Intensive care units (ICU)	(Ismail et al., 2023)	Vertical laminar airflow		7206	Standard k-ε
			Horizontal laminar airflow			
			Temperature controlled airflow			
Transportation Spaces	Aircraft cabin	(Rai & Chen, 2012)			85263	RNG k-ε
	Vehicle cabin	(Chang et al., 2023)			77778	SST k-ω/ RNG k-ε
Educational Spaces	Classroom	(Mirzaie et al., 2021)			6458	RNG k-ε
		(Pirouz et al., 2021)			7211	k-ε
	Lecture Room	(Arpino et al., 2023)			8205	SST k-ω
		(Lin et al., 2015)			369	RNG k-ε
Office		(Pirouz et al., 2021)			9059	k-ε
Restaurant		(Li et al., 2021)				RNG k-ε
Residential Spaces		(Bahramian et al., 2023)	Mixing ventilation	15	35000	RNG k-ε
		(Zong et al., 2023)	Displacement ventilation	0.51	27679	SST k-ω
		(Liu et al., 2023)	Make-up air organization from window		63272	Standard k-ε
		(Feng et al., 2020)			66667	LES
		(A. Zhang et al., 2019)			36765	Standard k-ε
		(Plana-Fattori et al., 2014)			15714	Standard k-ε
Museum		(Bakry et al., 2022)			18450	Realizable k–ε

4. Summary and Discussions

In conclusion, this review paper analyzed 25 research papers focusing on geometry and grid generation, ventilation strategies, and turbulence model selection in the context of indoor airflow simulations. This review investigated various geometries with varying dimensions and complexities, reflecting the diverse indoor environments. The tetrahedral mesh was the most frequently employed cell type among the different mesh types. The RNG k-ε model was the most used in studied papers regarding turbulence models. Furthermore, a comprehensive overview of the various ventilation strategies employed in each type of indoor environment is presented. This information provides valuable insights into the diverse approaches to ensure optimal air quality and circulation in various settings.

References

Aganovic, A., Steffensen, M., & Cao, G. (2019). CFD study of the air distribution and occupant draught sensation in a patient ward equipped with protected zone ventilation. *Building and Environment*, *162*. https://doi.org/10.1016/j.buildenv.2019.106279

Almhafdy, A., Korany, H. Z., AlSaleem, S. S., & Cao, S.-J. (2023). Airflow distribution in hospital isolation rooms with different ventilation and exhaust vent configurations. *Indoor and Built Environment*, 1420326X2311774. https://doi.org/10.1177/1420326X231177460

Arpino, F., Cortellessa, G., D'Alicandro, A. C., Grossi, G., Massarotti, N., & Mauro, A. (2023). CFD analysis of the air supply rate influence on the aerosol dispersion in a university lecture room. *Building and Environment*, *235*, 110257. https://doi.org/10.1016/j.buildenv.2023.110257

Bahramian, A., Mohammadi, M., & Ahmadi, G. (2023). Effect of indoor temperature on the velocity fields and airborne transmission of sneeze droplets: An experimental study and transient CFD modeling. *Science of the Total Environment*, *858*. https://doi.org/10.1016/j.scitotenv.2022.159444

Bakry, M. S., Hamdy, M., Mohamed, A., & Elsayed, K. (2022). Energy saving potential in open museum spaces: A comparative hygrothermal microclimates analysis. *Building and Environment*, *225*. https://doi.org/10.1016/j.buildenv.2022.109639

Chang, T. B., Lin, Y. S., & Hsu, Y. T. (2023). CFD simulations of effects of recirculation mode and fresh air mode on vehicle cabin indoor air quality. *Atmospheric Environment*, *293*. https://doi.org/10.1016/j.atmosenv.2022.119473

Feng, G., Bi, Y., Zhang, Y., Cai, Y., & Huang, K. (2020). Study on the motion law of aerosols produced by human respiration under the action of thermal plume of different intensities. *Sustainable Cities and Society*, *54*. https://doi.org/10.1016/j.scs.2019.101935

Ismail, Y. A., Eldosoky, M. A. A., Rashed, M. R., & Soliman, A. M. (2023). Numerical investigation of indoor air quality in health care facilities: A case study of an intensive care unit. *Journal of Building Engineering*, *68*. https://doi.org/10.1016/j.jobe.2023.106143

Kalliomaki, P., Koskela, H., Waris, M., & Wei-Tze Tang, J. (2020). *Assessing the risk to healthcare workers of hospital-acquired infection from patients infected with aerosol-transmissible pathogens Transmission of respiratory viruses View project Rhinovirus View project*. www.iosh.com/reducing-hospital-infections

Kohanski, M. A., Lo, L. J., & Waring, M. S. (2020). Review of indoor aerosol generation, transport, and control in the context of COVID-19. *International Forum of Allergy and Rhinology*, *10*(10), 1173–1179. https://doi.org/10.1002/alr.22661

Li, Y., Qian, H., Hang, J., Chen, X., Cheng, P., Ling, H., Wang, S., Liang, P., Li, J., Xiao, S., Wei, J., Liu, L., Cowling, B. J., & Kang, M. (2021). Probable airborne transmission of SARS-CoV-2 in a poorly ventilated restaurant. *Building and Environment*, *196*. https://doi.org/10.1016/j.buildenv.2021.107788

Lin, S., Tee, B. T., & Tan, C. F. (2015). Indoor airflow simulation inside lecture room: A CFD approach. *IOP Conference Series: Materials Science and Engineering*, *88*(1). https://doi.org/10.1088/1757-899X/88/1/012008

Liu, Y., Li, C., Ma, H., & Dong, J. (2023). Investigation on the indoor environment during a whole cooking process under constant make-up air organization in a Chinese-style residential kitchen. *Indoor and Built Environment*. https://doi.org/10.1177/1420326X231152554

Liu, Y., Long, Z., & Liu, W. (2022). A semi-empirical mesh strategy for CFD simulation of indoor airflow. *Indoor and Built Environment*, *31*(9), 2240–2256. https://doi.org/10.1177/1420326X221089825

Lu, Y., Oladokun, M., & Lin, Z. (2020). Reducing the exposure risk in hospital wards by applying stratum ventilation system. *Building and Environment*, *183*. https://doi.org/10.1016/j.buildenv.2020.107204

Marashian, S., Sadrizadeh, S., & Abouali, O. (2022). Modeling particle distribution in a ventilated room with modified discrete random walk methods. *International Journal of Ventilation*. https://doi.org/10.1080/14733315.2022.2143062

Mirzaie, M., Lakzian, E., Khan, A., Warkiani, M. E., Mahian, O., & Ahmadi, G. (2021). COVID-19 spread in a classroom equipped with partition – A CFD approach. *Journal of Hazardous Materials*, *420*. https://doi.org/10.1016/j.jhazmat.2021.126587

Morawska, L., Tang, J. W., Bahnfleth, W., Bluyssen, P. M., Boerstra, A., Buonanno, G., Cao, J., Dancer, S., Floto, A., Franchimon, F., Haworth, C., Hogeling, J., Isaxon, C., Jimenez, J. L., Kurnitski, J., Li, Y., Loomans, M., Marks, G., Marr, L. C., … Yao, M. (2020). How can airborne transmission of COVID-19 indoors be minimised? In *Environment International* (Vol. 142). Elsevier Ltd. https://doi.org/10.1016/j.envint.2020.105832

Pirouz, B., Palermo, S. A., Naghib, S. N., Mazzeo, D., Turco, M., & Piro, P. (2021). The role of hvac design and windows on the indoor airflow pattern and ach. *Sustainability (Switzerland)*, *13*(14). https://doi.org/10.3390/su13147931

Plana-Fattori, A., Trelea, I. C., Le Page, J. F., Souchon, I., Pollien, P., Ali, S., Ramaioli, M., Pionnier-Pineau, E., Hartmann, C., & Flick, D. (2014). A novel approach for studying the indoor dispersion of aroma through computational fluid dynamics. *Flavour and Fragrance Journal*, *29*(3), 143–156. https://doi.org/10.1002/ffj.3190

Rahman, M. N. Y., Razlan, Z. M., Izham, M., Omar, M. I., Zambri, N. A. A., Shahriman, A. B., Zunaidi, I., & Wan, W. K. (2018). A Test of Possibility on Relative Humidity Function in Minor Operation Theatre. *IOP*

Conference Series: Materials Science and Engineering, *429*(1). https://doi.org/10.1088/1757-899X/429/1/012089

Rai, A. C., & Chen, Q. (2012). Simulations of ozone distributions in an aircraft cabin using computational fluid dynamics. *Atmospheric Environment*, *54*, 348–357. https://doi.org/10.1016/j.atmosenv.2012.02.010

Sadeghian, P., Bi, Y., Cao, G., & Sadrizadeh, S. (2022). Reducing the risk of viral contamination during the coronavirus pandemic by using a protective curtain in the operating room. *Patient Safety in Surgery*, *16*(1). https://doi.org/10.1186/s13037-022-00332-x

Sadrizadeh, S., Tammelin, A., Ekolind, P., & Holmberg, S. (2014). Influence of staff number and internal constellation on surgical site infection in an operating room. *Particuology*, *13*(1), 42–51. https://doi.org/10.1016/j.partic.2013.10.006

Satheesan, M. K., Mui, K. W., & Wong, L. T. (2020). A numerical study of ventilation strategies for infection risk mitigation in general inpatient wards. *Building Simulation*, *13*(4), 887–896. https://doi.org/10.1007/s12273-020-0623-4

Śmiełowska, M., Marć, M., & Zabiegała, B. (2017). Indoor air quality in public utility environments—a review. *Environmental Science and Pollution Research*, *24*(12), 11166–11176. https://doi.org/10.1007/s11356-017-8567-7

Wang, C., Holmberg, S., & Sadrizadeh, S. (2018a). Numerical study of temperature-controlled airflow in comparison with turbulent mixing and laminar airflow for operating room ventilation. *Building and Environment*, *144*, 45–56. https://doi.org/10.1016/j.buildenv.2018.08.010

Xu, C., Liu, W., Luo, X., Huang, X., & Nielsen, P. V. (2022). Prediction and control of aerosol transmission of SARS-CoV-2 in ventilated context: from source to receptor. *Sustainable Cities and Society*, *76*. https://doi.org/10.1016/j.scs.2021.103416

Zhang, A., Zhen, Q., Zheng, C., Li, J., Zheng, Y., Du, Y., Huang, Q., & Zhang, Q. (2023). Assessing the impact of architectural and behavioral interventions for controlling indoor COVID-19 infection risk: An agent-based approach. *Journal of Building Engineering*, *74*. https://doi.org/10.1016/j.jobe.2023.106807

Zhang, B., Guo, G., Zhu, C., & Ji, Z. (2019). Transport of aerosol by coughing in an air-conditioned space. *Proceedings of the Thermal and Fluids Engineering Summer Conference*, *2019-April*, 1341–1353. https://doi.org/10.1615/TFEC2019.hbe.028480

Zong, J., Ai, Z., & Ma, G. (2023). Accurate evaluation of inhalation exposure based on CFD predicted concentration in the breathing zone towards personalized and smart control. *Journal of Building Engineering*, *71*. https://doi.org/10.1016/j.jobe.2023.106404

Mapping simulation optimization requirements for construction sites: A study in the heavy-duty vehicles industry

Abdulkarim Habbab[a,*], Anas Fattouh[b], Bobbie Frank[c], Koteshwar Chirumalla[b]
Markus Bohlin[b]

[a] *Mälardalen University / Volvo Construction Equipment, Box 325, 63105 Eskilstuna, Sweden,*
[b]*Mälardalen University, Box 325, 63105 Eskilstuna, Sweden,*
[c] *Volvo Construction Equipment, Box 325, 63105 Eskilstuna, Sweden*
*Corresponding author: abdulkarim.habbab@mdu.se

Abstract

The Construction and mining Industry comprises complex operations and interactions between various actors at different levels. Simulation has emerged as a valuable tool in this domain to better understand the site's behavior and optimize its operation. However, developing a simulation platform that can handle all the operations on the site is challenging due to the computational cost of the digital representation of reality along with the required accuracy level.

This paper aims at extracting and mapping the optimization requirements of construction sites at three main levels: site level, operational level and dynamics level. More precisely, this work seeks to define and map the most important requirements between these levels that ensure simulation credibility and reliability.

Based on interviews with experts in the domain, both from academia and industry, several key insights and recommendations emerged: at the site level, the layout and the key performance indicators, such as productivity, time, cost, number of machines and workers, need to be modeled and simulated. At the operational level, the simulation platform must include the main activities, such as loading, excavating, transporting and dumping. Moreover, the dynamics level should involve machine models and their interactions with the site's environment, such as earthmoving, drilling, excavating and blasting.

1. Introduction

In today's rapidly evolving digital landscape, digitization, as an essential pillar of Industry 4.0, is presented as a great enabler in the industry in terms of efficiency and productivity (Hermann et al., 2016). This is because it allows companies to leverage cutting-edge technologies, such as artificial intelligence, the Internet of Things (IoT), and machine learning, to organize operations and optimize resource utilization. In this context, simulation technology can be a great service to be digitized and presented to a broader range of end-users in the construction and mining industry (Tsai et al., 2016). Quarries, construction and mining sites are vital and dynamic work environments that involve harsh operations such as blasting, digging, excavating, and loading different kinds of materials. In order to optimize the workflow in these sites and maximize productivity while minimizing the cost and meeting environmental constraints, it is vitally important to have a well-designed management and control system in place at each level, site, operations and machines, with well-defined KPIs at each level.

At the site level, the main KPIs of the site related to productivity, time and cost calculations can be determined, which provide a comprehensive overview of the site's overall performance. At the operational level, we can define the main operations that will be ongoing in the selected site, including the type of equipment needed, the manpower required, and the safety protocols that need to be followed. Finally, and most importantly from a software engineering perspective, the machine dynamics level is where we set the proper parameters to represent the machines and facilities on the site, and that includes the type and specifications of used machines, and the modeling of machines' dynamic and their interaction with each other and with the surrounding environment.
By replicating real-world scenarios in a virtual environment, site simulation can be a great help to site designers, managers and engineers to plan, monitor, manage, and predict hazards and potential failures in their sites. However, the use of site simulation is often constrained by a significant bottleneck represented by the computational cost of modeling and simulating all actors operating on the

site at different levels. The main challenge is acquiring the required accuracy with the increased number of simulated machines, tasks, and facilities, creating a trade-off between accuracy and feasibility.

This work is motivated by two aspects: first, the current simulation tools are limited in their ability to provide a comprehensive site simulation due to their high computational cost (Guo & Zhang, 2022), and the huge amount of involved elements that can be in a site (Wickberg et al., 2022). Second, the decision-making process for designing and operating construction and mining sites is heavily based on experience rather than data-driven or standard insights.

This study represents an initial exploration in the field of full site modeling and simulation, with the intention of laying a foundation for further research. Subsequent studies will be conducted to expand upon and refine the findings presented herein, in order to contribute to the ongoing advancement of knowledge in this field. Moreover, future research endeavors will also aim to develop and test algorithms that can effectively address the problem explored in this study. These efforts will contribute to the advancement of the field and aid in the development of practical solutions for real-world applications.

The structure of this paper is as follows: Section 2 delves into a review of related literature, followed by an explanation of the research methodology in Section 3. The results are presented in detail in Section 4, and finally, a summary and discussion are provided in Section 5.

2. Related Work

Mapping the optimization requirements is defined as the process of understanding and identifying the key components, constraints, and variables involved in the optimization problem and representing them in a structured framework that can be used for analysis and solution finding.

Scientific software projects often overlook formal requirement engineering, as scientists may not see its benefits and lack knowledge in the field. A recent approach was to use natural language processing techniques, which achieved a high accuracy rate (Li et al., 2015). However, extracting these requirements in practical field is essential part of solving the problem and it requires a deep understanding of the involved processes and factors (Bashir et al., 2023).

In the following subsections, the evolutionary trajectory of earthwork and allocation optimization techniques will be presented, alongside an exploration of the approaches employed for addressing and resolving these optimization challenges.

2.1 Earthwork Optimization: Evolution and Categorization

In the context of construction site optimization requirements, there has been a growing interest in understanding the impact of earthwork optimization on mining and construction site operations. Starting from 1958 with Kantorovich and his proposal of a simple mathematical method to allocate materials to minimize the total cost of transportation of 1 m^3 of material from point A to B, up until 2021 where we start to see that it evolves to become multiple objectives optimization problem (Fernandes et al., 2022).

Delving into the body of literature through a meticulous review conducted by Fernandes et al. (2022), distinct categories within the domain of earthwork optimization emerge, such as:

2.1.1. Earth Allocation Planning:

This category is primarily dedicated to the minimization of costs, often constrained by financial considerations encompassing excavation, hauling, and compaction expenses.

2.1.2. Equipment Fleet Planning:

In this segment, studies gravitate towards multi-objective strategies that extend beyond cost optimization, encompassing concerns related to fuel consumption and emissions reduction.

2.1.3. Earthmoving Operation Routing and Scheduling:

Researchers in this area are committed to optimizing costs, time, and transportation distances through judicious routing and scheduling of earthmoving operations.

Despite the strides taken in earthwork optimization, a critical observation arising from the reviewed studies is the often-peripheral treatment of environmental effects linked to earthmoving operations within optimization objectives. This vital aspect highlights a potential gap in current research priorities.

2.2 Methodological Approaches to Earth Allocation Optimization

Numerous methodologies have been employed to tackle the intricate challenge of earth allocation optimization, such as:

2.2.1 *Genetic Algorithms, Linear Programming (LP), and Mixed-Integer LP:* These traditional techniques serve as the cornerstone. Yet, challenges like discrete domain variables and nonlinear relationships can impede their effectiveness (Burdett & Kozan, 2014; de Lima et al., 2020; Fattouh et al., 2021; Lim et al., 2017; Montaser et al., 2012).

2.2.2 Hybrid Approaches: A notable contemporary trend involves the fusion of algorithms, such as genetic algorithms and LP or combinations like tabu search and simulated annealing (Burdett & Kozan, 2014; de Lima et al., 2020; Fattouh et al., 2021; Lim et al., 2017; Montaser et al., 2012).

2.2.3 Fleet Planning: Evolutionary techniques assume significance in this domain, often synergizing with GPS and GIS methodologies to enhance precision in productivity estimations (Montaser et al., 2012).

It's imperative to acknowledge the limitations intrinsic to each of these algorithms, arising from factors such as discrete domain variables or nonlinear relationships.

2.3 Innovations in Site Layout Planning Optimization

Turning attention to site layout planning, research predominantly leans towards heuristic or metaheuristic models, driven by mathematical constraints in optimization software and the inherently multifaceted nature of site layouts, featuring diverse objectives, laborer and equipment pathways, and facility configurations (Kaveh & Vazirinia, 2018). An underlying premise in optimization algorithms is the alignment between their search paths and the search capability they provide (Xu et al., 2020).

3. Methodology

This research was performed using the case study procedures (Patton, 1990), as the study was designed based on a certain number of workshops and interviews with experts in different levels of the studied case. The selection of the interviewed people was based on their relevance to the involved project, as the selected experts were chosen to cover different operational levels from business to operations and finally technical software modeling and simulation. Table 1 shows the details of the conducted interviews and workshops.

Table 2 shows the participants' designation. The workshops were conducted on a regular basis to discuss the best practices for general construction and mining site optimization solutions.

The interview questions were set to extract the main site design considerations and map them downwards to achieve the most ideal machines, materials and facilities digital representation. The questions were designed to distinguish different site levels (site, operations and machine dynamics), this distinguish comes from the software design needs for more dynamic architectural design of the suggested framework (Fattouh, 2022), as one of its main features is the ability to be expanded always to include new machines, material types and different kind of facilities. The goal of these questions was to get the boundaries on each site level that are essential to keep the site in ideal operation mode, how to achieve that currently and to explore the possibilities of mistakes and problems that can happen on the site and map all this information together to help to build better models of the machines, material their interaction and finally the whole site in general.

We interviewed one site manager, one business development engineer and several modeling and simulation experts. The interviews were conducted online with an approximate timing of 2-3 hours each, while the workshops' timing was approximately 6 hours each.

Table 1: Conducted interviews and workshops.

Description	Interview	Workshop
Number of conducted sessions	6	5
Time/session	2-3h	6h
Total Number of Experts	5	6-12

Table 2: The participants' designation.

Title	Functional Organization	Years of Experience
Simulation Engineer	Product Platform / Wheel Loader	15
Calculation Engineer	Virtual Product Development	17
Research Owner	Emerging Technologies	16
Senior Lecturer	Product Realization Division	5
Business Development Engineer	Stone materials Department	7
Construction Site Manager	Stone materials Department	19

4. Results

The objective of this research is to establish a comprehensive framework for defining and categorizing optimization requirements pertaining to quarries, construction, and mining sites, with the

aim of achieving a practical and dependable approach to site modeling and simulation.

Findings from interviews and workshops, particularly with industry experts, indicated that a significant number of design decisions and considerations continue to be influenced by experiential knowledge, business objectives, and regulatory constraints, as these factors define the realistic boundaries within which site operations are conducted
Figure 1.

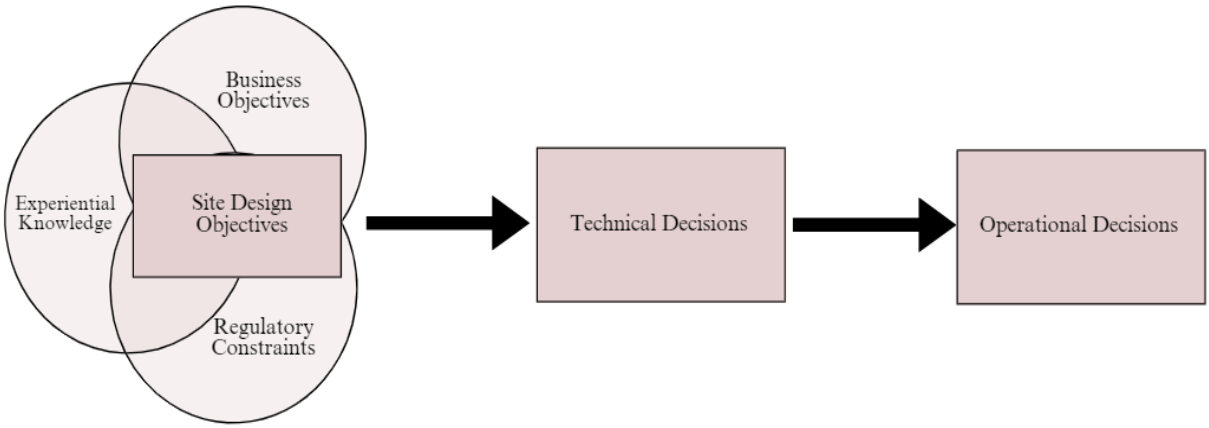


Figure 1 The factors influencing the decision-making processes.

Additionally, from a software perspective, simulating complete construction and mining sites involves significant computational expenses, specifically when considering diverse machinery types, materials, and the presence of static and dynamic objects. The behavior, interactions, and movements of machinery like excavators, bulldozers, cranes, and trucks require complex algorithms. Material simulations encompass forces, deformation, and flow for various materials such as soil, rock, aggregates, and ores. Simulating interactions and collisions between static objects (structures, terrain) and dynamic objects (vehicles, personnel) necessitate sophisticated algorithms and precise modeling of dynamics, kinematics, and spatial relationships.
Consequently, it is imperative to carefully examine the design considerations pertaining to models representing sites, machinery, materials, and machine-material interactions, in order to mitigate the exponential escalation of system complexity.
The resulting requirements and recommendation for simulation optimization can be summarized as follows.

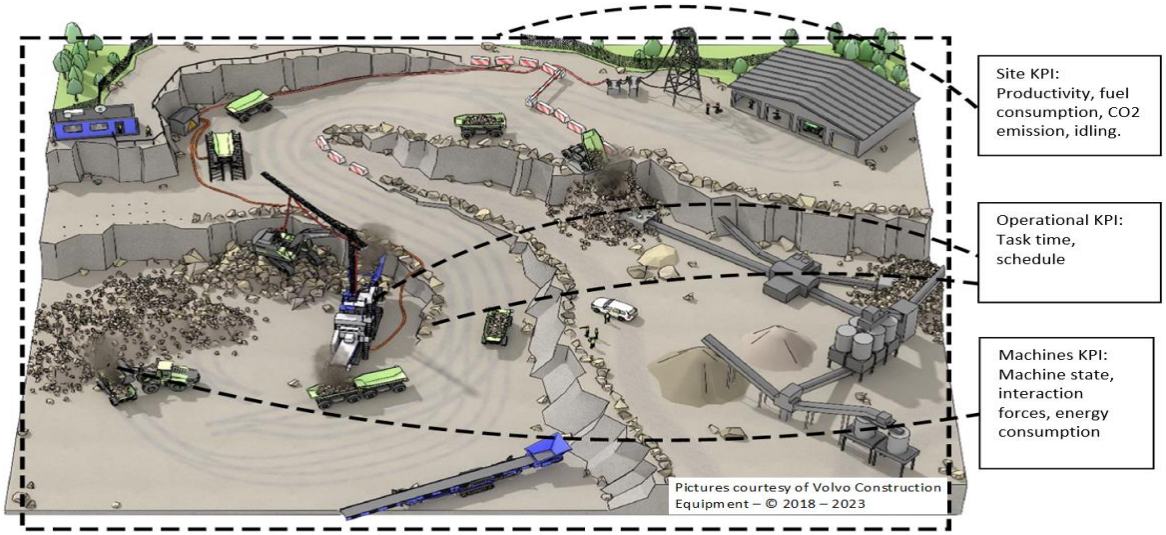


Figure 2: Important KPI at each level in a quarry site.

4.1. General KPI's of the Site

There are several factors that need to be considered when evaluating a site for a project (Figure *2*). These include but are not limited to productivity, cost, time, energy consumption, environmental requirements, and governmental permissions. It is important to mention that these factors are highly dynamic throughout the lifespan of the site and adapting to these factors together could be challenging. Governmental permissions are a critical consideration that can affect various aspects of a project. These permissions dictate working hours, safety measures, noise levels, and environmental constraints. Understanding and complying with these regulations is crucial for legal compliance and avoiding potential penalties or delays. Additionally, governmental permissions can align with business objectives, ensuring that the project meets all necessary regulations while achieving its desired outcomes.

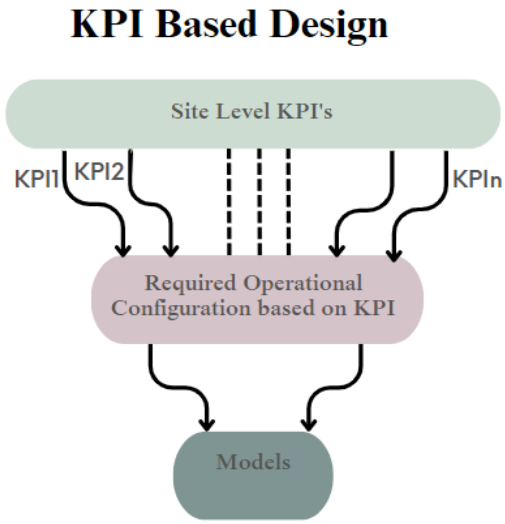


Figure 3 KPI-Based design and models selection

Changes in project goals, market conditions, or regulatory frameworks can introduce new challenges and considerations. Therefore, it is essential from a software perspective to have adaptable characteristics and decision-making processes that can accommodate these evolving factors Figure 3. These characteristics include:

4.1.1. Flexibility and Customizability: to accommodate any change in the targeted KPI or the nature of the site itself based on the need of the project like new routes, increased or decreased number of machines and operators, different kinds of equipment... etc.

4.1.2. Integration: to have a successful, meaningful, and smooth simulation the software must have the ability to integrate and interact with other tools and software.

4.1.3. User Interface: to give the user the ability to set up the simulation in an easy and meaningful manner a user-friendly interface is essential in such an application.

4.2. Design based on Task

Designing the models of machines, material and their interaction in such applications is the most challenging part as this problem can easily grow in complexity to a very high level due to the huge number of parameters needed to represent the site and the activities going on in an accurate way, and due to the nature of these activities that has a huge complexity level in its nature as it is heavily dependable on previous state of the model in each task independently. That is why these models should be designed in a modular flexible way where certain parts of the model can be enabled or disabled based on the nature of the task. The main tasks that we will discuss are the loading and dumping task and the moving task.

During the loading and dumping task, the main factors to estimate the fuel consumption and subsequently the other relevant KPIs like productivity and CO2 emission are the forces generated from the bucket-material interaction. The machines mainly articulated wheel loaders, or excavators, will be in a relatively stationary state, where the bucket and hydraulic systems are mainly the active systems to perform the loading and dumping processes (Filla, 2005; Frank et al., 2018). Modeling and simulating these processes represent one of the main bottlenecks in the site's simulation. This interaction problem can be approached through various methods, including numerical-based approaches, trajectory-based approaches, and machine learning techniques.

The numerical-based approach in site simulation involves approximating the interaction between the machine's bucket and the material pile as two sides of an equation. The goal is to establish a relationship between the parameters representing the forces involved in the interaction. Although this approach shows feasibility and promise, it encounters a significant challenge when applied to large-scale sites with numerous machines in operation.

Trajectory-based approaches aim to estimate the state parameters of the interaction between the bucket and the material. This method relies on modeling the interaction based on the previous states of the bucket along its trajectory during the excavation process, Notably, the distinct pathways followed by the bucket during excavation lead to the generation of differing forces inherent to this

excavation procedure, Figure 4 shows the possible trajectories that could be taken in the excavation process. However, this approach introduces a complexity problem akin to the Travelling Salesman Problem (TSP), as the estimation of forces changes with each new penetration into the material pile. Consequently, this dynamic estimation challenge poses a significant obstacle to overcome in the trajectory-based approach.

Machine learning techniques have become increasingly popular for addressing the interaction problem in site simulation (Egli et al., 2022). By training on large datasets, machine learning algorithms can recognize patterns, learn from past simulations, and predict system behavior. This approach is particularly valuable in complex and dynamic environments where traditional analytical models may be limited. Through machine learning, simulation models can continuously improve and adapt, leading to enhanced accuracy and predictive capabilities.

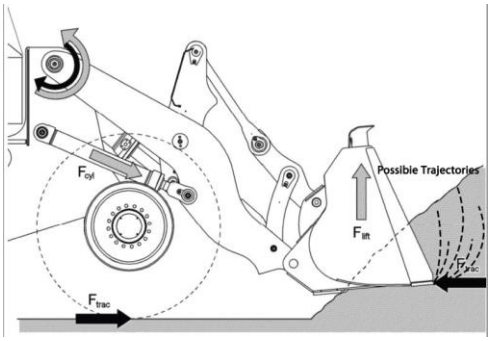


Figure 4: Trajectory-based approach - possible trajectories that can be taken through the excavation process. modified from [(Filla, 2005)].

During the moving task, the machine model exhibits a reduced level of complexity as compared to the model employed in the prior task. Nevertheless, several factors exist that could potentially contribute to increased system complexity. The impetus to minimize this complexity stems from the overarching goal of attaining a comprehensive view of the site simulation environment in this case study. Our findings suggest that a few key parameters, namely the road profile, power train, and axle load, have significant impacts on the energy management calculations during the moving task, Therefore, these parameters should be considered the primary components in the task's model design and subjected to thorough analysis.

The optimization problem faced by specialists in the construction, mining and quarries industries encompasses multiple dimensions and is evident both during the initial planning phase of opening a new site and in daily operations. The inherent limitations of humans make it challenging to effectively solve multidimensional problems, resulting in suboptimal solutions. However, these challenges highlight the need for alternative approaches.

Looking toward the future, it is anticipated that the number of dimensions involved in site optimization will increase further. For example, achieving zero emissions may require the use of battery electric machines, which necessitate more frequent and longer charging times compared to refueled equipment. Additionally, automation is expected to play a significant role, especially in sites with mixed traffic and partial automation, necessitating a different design approach (Frank, 2019; Wickberg et al., 2022).

5. Summary and Discussions

This research aimed to extract and map the simulation optimization requirements among quarries, construction, and mining sites. The study revealed that site design and management decisions are highly influenced by experiential knowledge, business objectives and regulatory constraints. These factors along with the complicated nature of the daily site operations put the site into a suboptimal state and this problem can be solved by simulation. Moreover, simulating complete sites presents its own set of challenges. The dynamic nature of site processes, coupled with the presence of diverse machinery, materials, and static and dynamic objects, introduces significant computational demands.

Different approaches were discussed with experts from the industry to recommend suitable methods to build and simulate sites and optimize the simulation. The study highlighted the importance of adaptable software that can accommodate the dynamic nature of sites, modular design approaches that allow for flexibility and customization, and the consideration of key parameters in the simulation process.

Overall, the research offered valuable insights for effective site simulation optimization. The next step will be to investigate the best machine learning approach to solve the complexity problem and compare that to the performance of numerical approaches.

Acknowledgment

This work is funded by the KK-Stiftelsen via the project (IndTech+) – Industrial Technology Graduate School, grant number 20200132 01 H, and by the Vinnova FFI project TRUST-SOS - TRUSTed Site Optimisation Solutions, grant number 2021-02551.

References

Bashir, S., Abbas, M., Saadatmand, M., Enoiu, E. P., Bohlin, M., & Lindberg, P. (2023, 2023//). Requirement or Not, That is the Question: A

Case from the Railway Industry. Requirements Engineering: Foundation for Software Quality, Cham.

Burdett, R., & Kozan, E. (2014). An integrated approach for earthwork allocation, sequencing and routing. *European Journal of Operational Research*, *238*, 741–759. https://doi.org/10.1016/j.ejor.2014.04.036

de Lima, R., Nobre Júnior, E., & Pinheiro Santos Fernandes, P. (2020). Optimization of Earthmoving Operations Planning: A Novel Approach Considering Interferences. *Journal of Engineering Project and Production Management*, *11*, 158-168. https://doi.org/10.2478/jeppm-2021-0016

Egli, P., Gaschen, D., Kerscher, S., Jud, D., & Hutter, M. (2022). Soil-Adaptive Excavation Using Reinforcement Learning. *IEEE Robotics and Automation Letters*, *7*(4), 9778-9785. https://doi.org/10.1109/LRA.2022.3189834

Fattouh, A. (2022, 9-11 Dec. 2022). An Integrated Simulation Framework for Construction Site Operations. 2022 IEEE 1st Industrial Electronics Society Annual On-Line Conference (ONCON),

Fattouh, A., Bohlin, M., & Sundmark, D. (2021, 23-26 April 2021). A Real-Time Optimization Model for Production Planning in Quarry Sites. 2021 IEEE 8th International Conference on Industrial Engineering and Applications (ICIEA),

Fernandes, P. G. P. S., Júnior, E. F. N., & Prata, B. d. A. (2022). Optimization of earthworks planning: a systematic mapping study. *Canadian Journal of Civil Engineering*, *49*(12), 1781-1795. https://doi.org/10.1139/cjce-2022-0185

Filla, R. (2005). *Operator and machine models for dynamic simulation of construction machinery* Institutionen för konstruktions-och produktionsteknik].

Frank, B., Kleinert, J., & Filla, R. (2018). Optimal control of wheel loader actuators in gravel applications. *Automation in Construction*, *91*, 1-14. https://doi.org/https://doi.org/10.1016/j.autcon.2018.03.005

Frank, M. (2019). *A Step Towards the Design of Collaborative Autonomous Machines - A Study on Construction and Mining Equipment*

Guo, K., & Zhang, L. (2022). Multi-objective optimization for improved project management: Current status and future directions. *Automation in Construction*, *139*. https://doi.org/10.1016/j.autcon.2022.104256

Hawarneh, A. A., Bendak, S., & Ghanim, F. (2021). Construction site layout planning problem: Past, present and future. *Expert Systems with Applications*, *168*. https://doi.org/10.1016/j.eswa.2020.114247

Hermann, M., Pentek, T., & Otto, B. (2016). *Design Principles for Industrie 4.0 Scenarios* 2016 49th Hawaii International Conference on System Sciences (HICSS),

Kaveh, A., & Vazirinia, Y. (2018). Construction Site Layout Planning Problem Using Metaheuristic Algorithms: A Comparative Study. *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, *43*, 105-115.

Li, Y., Guzman, E., Tsiamoura, K., Schneider, F., & Bruegge, B. (2015). Automated Requirements Extraction for Scientific Software. *Procedia Computer Science*, *51*, 582-591. https://doi.org/https://doi.org/10.1016/j.procs.2015.05.326

Lim, A., Rodrigues, B., & Zhang, J. (2017). Tabu search embedded simulated annealing for the shortest route cut and fill problem. *Journal of the Operational Research Society*, *56*(7), 816-824. https://doi.org/10.1057/palgrave.jors.2601900

Montaser, A., Bakry, I., Alshibani, A., & Moselhi, O. (2012). Estimating productivity of earthmoving operations using spatial technologies1This paper is one of a selection of papers in this Special Issue on Construction Engineering and Management. *Canadian Journal of Civil Engineering*, *39*(9), 1072-1082. https://doi.org/10.1139/l2012-059

Optimization of earthworks planning a systematic mapping study. *Canadian Journal of Civil Engineering*.

Patton, M. Q. (1990). *Qualitative evaluation and research methods, 2nd ed*. Sage Publications, Inc.

Tsai, W. T., Zhibin, C., Xiao, W., Paul, R., Qian, H., & Xin, S. (2016). Modeling and Simulation in Service-Oriented Software Development. *Simulation*, *83*(1), 7-32. https://doi.org/10.1177/0037549707079221

Wickberg, P., Fattouh, A., Afshar, S., Sjöberg, J., & Bohlin, M. (2022, 27-29 Dec. 2022). Dynamic Maps Requirements for Autonomous Navigation on Construction Sites. 2022 5th International Conference on Communications, Signal Processing, and their Applications (ICCSPA),

Xu, M., Mei, Z., Luo, S., & Tan, Y. (2020). Optimization algorithms for construction site layout planning: a systematic literature review. *Engineering, Construction and Architectural Management*, *27*(8), 1913-1938. https://doi.org/10.1108/ecam-08-2019-0457

AN EMBEDDED INDUSTRIAL CONTROL SYSTEM FRAMEWORK FOR MODEL PREDICTIVE CONTROL OF A DISTRICT HEAT SUBSTATION

Joakim Örneskans [a,*], Konstantinos Kyprianidis [b], Stavros Vouros [c], Gunnar Bengtsson [d]

[a] *Mälardalen University,* [b] *Mälardalen University,* [c] *Mälardalen University,* [d] *First Control Systems AB*
joakim.orneskans@mdu.se

Abstract

In this paper we present a standard platform XC05 for an Edge Controller based on an Industrial Control System, where functions made in Modelica and Python can be run as an integrated part of an automation system. We demonstrate how the platform is used to run a complex Model Predictive Control (MPC) strategy to optimize indoor heating in a residential building. MPC strategies have been increasingly popular due to their ability to handle nonlinear dynamics with constraints and multi-objective optimization. Since industrial control systems are real-time based, consideration must also be taken to running security and the real-time characteristics and timing of the overall system solution. We also show that heavy calculation, protected by the industrial control system operative, can run safely together within fast automation using standard electronics. The controlled variable in the MPC strategy is the supply water temperature (Space heating), and the objective is to keep the indoor temperature at a predefined setpoint despite variations in outdoor weather conditions by using local measurements and weather forecasts from the Swedish weather service SMHI. The model used in the MPC is trained automatically with real-time data during running. We describe the controller architecture and briefly the model predictive control algorithm, analyze the overall system performance regarding safety and real-time characteristics. The proposed model predictive control application showed stable operation and expected real-time characteristics during operation. Furthermore, a reduction in indoor temperature deviations was achieved.

1. Introduction

The XC05 automation platform was developed as an activity in the DISTRHEAT research project where the aim of the project is to demonstrate and test in real operating environment (MPC) applied to District Heating and Cooling networks. The automation platform is in fact an “Edge controller” as expressed in the standard Industry 4.0 i.e., a highly intelligent unit interfacing the process or machine. Such controllers are expected to replace conventional PLC systems in the future. The details of the software design are described below.

In the DISTRHEAT project we applied our ideas to Model predictive control (MPC) for optimizing the indoor temperature in residential buildings. Model predictive control (MPC) is an optimal control technique where the control actions minimize a cost function over a finite specified time horizon.[1]

Using MPC for controlling different processes in the heating sector has proven to be highly successful as shown in [6] [7] [8] and [9]. To implement an MPC in a conventional control system (PLC) has been done [2] but requires deep understanding of the underlying PLC language and is often limited to special use-cases. The novelty of the proposed solution is using well known languages and tools which interact in a safe way with the time-critical functions of the PLC software,

The research project includes all the basic steps of development such as control methods, simulations, and installation in a physical process.

The automation platform was developed to cover all these aspects. It includes an industrial control system (ICS) which is integrated with commonly used research tools like Python scripts and Modelica simulations. In XC05 we introduce a new standard how to integrate research tools with an ICS system. For this purpose, we developed a graphical tool, FirstGraph where the objects in the ICS and the extended Python and Modelica libraries can be connected in a simple way. The tool is an extension of what we previously have used in the ICS and is understood by any process engineer. At each time step, the MPC controller receives or estimates the current state of the process being controlled. It then calculates the sequence of control actions that minimizes the cost over a specified time horizon by solving a constrained optimization problem that relies on an internal process model and the current system state. The controller then applies the first

computed control action for the building and the procedure is then repeated each time instance. The internal model in the MPC algorithm is automatically updated from real-time data during periodical learning periods which are run in parallel to the control actions. Data are assembled and stored by the ICS part. The temperatures in the building are read from a local PLC via a standard Modbus TCP communication line and the weather forecast are read periodically from the Swedish Meteorological and Hydrological Institute (SMHI).

This MPC application was a good test of the XC05 platform since it consists of both heavy calculations in the MPC algorithm and fast real time data handling. Moreover, the platform had to integrate MPC software made by other researchers with the ICS functions and safety supervision. Before it was connected to the building, the MPC application was run in a real-time simulation inside the XC05 to verify the operation, capacity and safety in the simple standard electronics used (Raspberry 4). Linux with the RT_PREEMPT patch is chosen as the operative system because of its stable real-time performance [3][4][5]. The MPC installation has now been in operation for more than five months without a single operation failure.

2.Methodology

The basic task was to integrate an ICS system with functions created in Modelica and Python. The new platform is based on the same principles that were used in the ICS system, that is used as a base for the platform.

2.1. Industrial control system

We choose our own ICS which has been used in many industrial installations and therefore is very safe and contains all the functions needed in an automation task including our own version of adaptive control. It is based on the self-tuning concept and has been awarded by the IEEE. The ICS is then transformed to the Linux operative with necessary changes to preserve the original properties of the ICS platform.

2.2. Python scripts

The python scripts are made in the normal way in a standard PC and then loaded into the XC05 platform where they are stored in a special library dedicated to Python scripts. They must be supplied with simple standard input/output functions to interface the ICS.

2.3. Modelica simulation

The Modelica simulation is designed in the normal way on a standard PC and then loaded into the XC05 where it is compiled locally and stored in a special library dedicated to Modelica models.

2.4. Supervision and safety

The Python scripts and the Modelica models are external software which may contain errors. Therefore, they must be supervised and disconnected in case of malfunction or a software error to prevent disturbance in the automation system. The ICS part contains by itself safety protection developed for very sensitive processes in the steel and energy areas.

2.5. Functional integration

All functions in the platform are regarded as “modules” or “objects” with specified inputs and outputs according to the same standard as is used in the ICS system. This means that much of the software developed for the ICS can be reused in this case.

2.6 MPC algorithm

The core focus of this paper is the ICS edge controller, therefore only a brief description will be provided for the MPC kernel that runs inside. The MPC is a Python-based optimizer that is employing a machine learning model for building heat demands and thermal comfort. The objectives of the optimizer comprise multiple indices related to energy performance of the heating system or indoor temperature in the apartments.

Anyone familiar with control theory knows that PID technique developed for more than 50 years ago is far too limiting in this case since its internal structure is limited, not prepared for predictive control, and does not support feedforward. According to the internal model principle, any efficient control algorithm must contain an internal model of the process and its disturbances. A candidate would be a general adaptive controller based on the self-tuning principle, which is in fact an adaptive MPC based on a linear dynamical model. Such a regulator is available in the ICS part of XC05 and has been used in many installations. In this case, however, an MPC controller was favored since it is more general, can handle nonlinear physical models and the optimization algorithm can be adjusted to the specific case. Such a solution has also a higher development potential in the future. The drawback is the computational burden caused by the optimization and model training, but we have demonstrated it is quite feasible within the XC05Edge controller.

2.7 Experimental study

An experimental study was conducted in a residential building. The MPC algorithm was developed and customized for the building's district heat substation. It used real-time data from the local PLC, such as temperature and outdoor weather conditions, to optimize the operation of the system. The MPC algorithm determined the optimal setpoints for the forward temperature based on keeping the indoor temperature at a fixed setpoint. To evaluate the performance of the system, data were collected at one- and sixty-minutes intervals throughout the test period. Descriptive statistics, such as mean value and standard deviations were calculated before and after installation of the system.

3. Edge computing

The automation platform described in this paper is an "Edge Controller" as expressed in Industry 4.0. Such controllers are expected to replace conventional PLC systems in the future.

An important issue is to spread research results to ordinary process engineers. This means that the way the engineer interfaces the new technology should be according to the standards used in common automation.

Our previously developed tool FirstGraph for the ICS, which has been used by engineers for several years, has been extended to the generalized XC05 platform. It provides a standard format where all functions are regarded as "objects" which are connected to each other. The platform then unifies the automation functions in ICS with external functions created in Modelica or Python.

The FirstGraph project tree has been extended with two new libraries Modelica and Python.

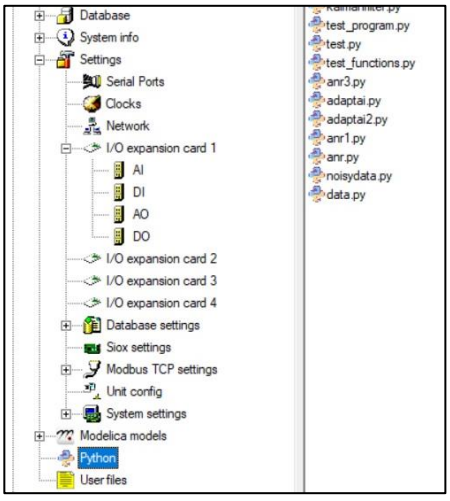


Figure 1: The XC05 graphical programming interface for engineers.

The process engineer selects elements from the extended libraries and connects them graphically to the automation system as is illustrated in figure 7. This can be done without stopping the running control task. Note that a program change will be active within about 0.5 msec after being loaded since it is handled locally by the operative.

If there is an MPC function or a Modelica function loaded to the library, the process engineer may directly use it in the automation system which creates a direct link between researchers and users. This was an important factor in this development.

4.Proposed architecture

4.1. Embedded system Architecture

To make the entire system portable to different hardware architectures all software are developed in C99 targeting different Linux distributions. It is also possible to port the framework to other operating systems by using a simple API to interface the new operating system.

The system consists of four main modules and a supervisor where all are strictly prioritized based on their different functionalities.

These modules are:

- Main controller task (ICS)
- Simulation executor
- Supervisor task
- Communication task
- Python executor

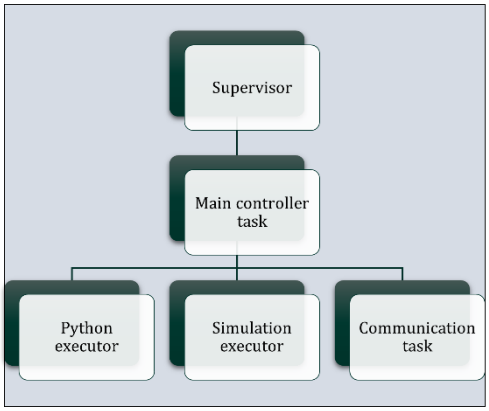


Figure 2: Embedded system architecture

The main controller task (ICS) handles all PLC functionality and executes all user applications. These applications could be run on eight different priority levels depending on their purpose. It's also responsible for creating the different subtasks and supervising these.

The simulation executor is responsible for running simulation of user-defined models created in the OpenModelica language and communication with the main controller task by using shared memory.

All external communication is handled by the communication task which supports a variety of

industrial protocols like Modbus TCP/IP, Modbus RTU and others. The Python executor handles the execution of user defined python scripts and communicates with the main controller task by using POSIX queues. Each parent task in the systems supervises and receives errors from subtasks. All errors are forwarded to the supervisor task which takes different actions depending on the severity of the error. More detailed information about the principles in the XC05 platform [10].

4.2 ICS task

The threads in the main controller task are strictly prioritized with respect to their functionality. Application level indicates nine different levels (priority 67 to 75, 75 being the highest) where user-applications could be executed at different priorities. The system is strictly event based and the threads are only woken when some external- or timer event occurs. The main controller task has the highest thread priorities in the system except for the Supervisor task main thread. The communication task main thread will inherit the same priority as the thread handling the communication in the main controller task (COMM1 and COMM2). To minimize latency times, all threads which belong to the main controller are directed to a single CPU-core.

Name	Type	RT-Priority (80 highest)
MAIN	System thread	80
TIMER	System thread	79
PCHH	System thread	76
APPLEVEL1	Application thread level 1	75
APPLEVEL2	Application thread level 2	74
APPLEVEL3	Application thread level 3	73
APPLEVEL4	Application thread level 4	72
APPLEVEL5	Application thread level 5	71
APPLEVEL6	Application thread level 6	70
APPLEVEL7	Application thread level 7	69
APPLEVEL8	Application thread level 8	68
APPLEVEL9	Application thread level 9	67
COMM1	System thread	65
COMM2	System thread	64
UDPIN	System thread	63
MTCP	System thread	62
MDUPOUT	System thread	60
MTCPS	System thread	58
PYTHONSUPERVISOR	System thread	47
MODELICASUPERVISOR	System thread	46
TERMIN	System thread	44
TERMOUT	System thread	42
CYCLIC	System thread	38
FCOMMIN	System thread	36
FCOMMOUT	System thread	34
BACKCALC_HI	System thread	32
PCHL	System thread	30
OPCOM	System thread	28
MODELCACOMP	System thread	26
BACKCALC_LO	System thread	24

Figure 3: Threads and priorities main controller task

The applications written in the ICS by the user are executed on APPLEVEL1 – 9 which have among the highest priorities in the system.

4.3 Simulation executor

The simulation executor is responsible for executing code written in the Modelica language which is used in various academic institutes as well in industry. In order to accomplish this the complete OpenModelica package is preinstalled on the system. This includes the OpenModelica compiler, the OpenModelica simulator and a large library of predefined objects.
The simulation executor consists of two parts, the Modelica compiler and the Modelica executor.

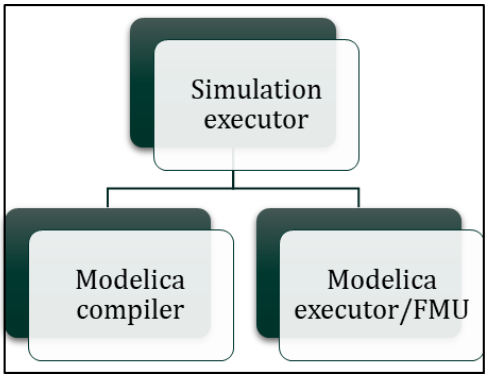


Figure 4: The simulation executor

The Modelica executor compiles the Modelica code defined by the user into a shared library according to the Function Mock-up Interface (FMI) into a Functional Mock-up Unit (FMU). The main controller task communicates with the simulation executor using shared memory protected by POSIX semaphores.

4.4. Python executor

The application thread of the main controller task is responsible for creating each corresponding subtask to the Python code that should be executed. Each subtask is then supervised from python supervision thread in the main controller.

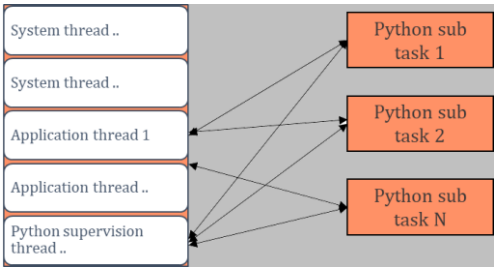


Figure 5: Python sub task creation

In order to support different scenarios of how the python code should be executed, two different modes are introduced, synchronous and asynchronous. In synchronous mode each subtask is created at the same priority level as the calling application thread which waits for an answer or a

user defined timeout before it continues execution. In asynchronous mode each subtask is created on priority level 0 (background) and the calling thread immediately continues execution and checks periodically if an answer has been received. A fixed limit of how many sub tasks that can be created has been set in the system due to system limitations.
The python code is called by an interface function defined in the python script by the user. The function can have an arbitrary name but must be defined with a specific number of arguments that corresponds to the executing block in the ICS. The function consists of a list of four which contains float, integer, Boolean and text values. The last three arguments are prepared to be returned by the python script and contains real, integers and Booleans.
The data exchange between the ICS and python executor is handled by POSIX message queues. Each sample, set by the user in the ICS, the corresponding ICS block checks if the python subtask is Idle. If the Python subtask is idle a message is sent to the message queue with the last values from the ICS. The Python subtask then receives the values, executes the script and returns the result to the ICS. Depending on if the python sub task is created in synchronous or asynchronous mode the corresponding block in the ICS waits for the answer or continues execution immediately.
The errors in the python script are handled by the corresponding block in the ICS. Any syntax error in the python script causes the python sub task to end execution.

4.5. Supervisor

The main functionality of the supervisor is to monitor all other sub tasks in the system. The supervisor has the highest priority in the system and is able to take predefined actions depending on if an error state exists in some of the sub tasks. This could be, for example, a controlled shutdown of all system tasks, restart of a certain sub task or a complete halt of the system. Each task communicates with the supervisor task using a shared memory which contains the current state of the specific task.

5. Test installation in a residential building

Figure 6: Test installation site

The experimental setup is performed at a district heat substation in a residential building in an urban area. The experiment is conducted in the heat season to get reasonable data for the MPC algorithm.
Hardware and Technical Stack: The framework runs on Linux (Debian 10) on a Raspberry Pi4. The detailed system setup and technical specifications to run the experiment are presented in Table 1.

Table: 1.

Parameters	Values
OS	ARM based Raspbian
OS Name	Debian 10
OS Version	Linux 5.11 RT-PREEMPT
Processor	Broadcom BCM2711 SoC with a 1.5 GHz 64-bit quad-core ARM Cortex-A72 processor
System RAM	2GB

We implement and evaluate the MPC algorithm using the Scikit-learn2 library, SciPy library and NumPy library for Python3. The data from different sensors are fetched from the local PLC by using Modbus TCP/IP with a time resolution of 50 milliseconds. The prediction of the local outdoor temperature for the next 10 hours is fetched from the Swedish weather service using SMHI API. The MPC algorithm then calculates the setpoint for each hour and writes it back to the local PLC by using Modbus TCP/IP. The data is saved locally with one- and sixty minutes time resolution for further analysis and to be processed by the MPC algorithm. The internal sample time of the ICS is set 10 milliseconds for each application task to register the real-time behavior of the overall system. Timestamps are recorded for each sample interval to be used for further analysis.

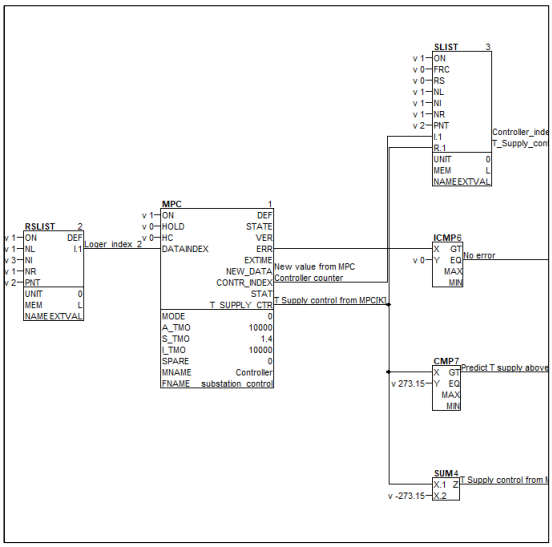


Figure 7: Application program in the ICS

6. Results of the test installation

The data gathering started at the end of January and the controller was turned on February 6th.

The complete system showed stable real-time characteristics during the whole test period regardless of the heavy background load due to the MPC computations. The ICS system functions were performed within the time intervals set by the user, which in this case was a sampling time of 10 milliseconds without any disturbances.

The Python-implemented MPC algorithm was executed as expected on the proposed platform and without interfering with the ICS system's functionality. This includes updates of python software, application changes including adding new applications programs and functions. When comparing the time required for a cycle's calculations, these were performed on a normal PC with a time consumption of approximately 4 minutes and on the industrial control system, the time consumption was approximately 8 minutes. This depends on hardware capabilities and process architecture.

The MPC algorithm performed very well and was stable throughout the whole test period. The purpose was to keep the indoor temperature at a predefined set-point, in this case 21.5°C and minimize the variations. The controlled variable was the supply temperature (space heating) which was set to the local PLC at site. As seen in figure 8 the indoor temperature was reduced from a mean value of 22.5°C to 21.5°C about one day after the controller was switched on. The variations were about +-0.2K.

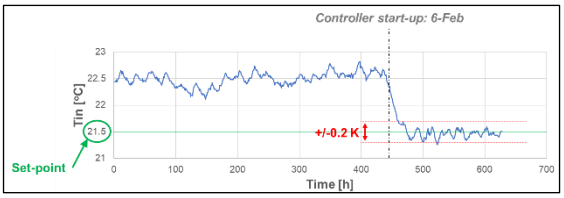


Figure 8: MPC performance

The response in temperature can of course be made much faster, but that would not have been convenient for inhabitants in the building. In this case, the control action is deliberately limited to achieve an "acceptable change in temperature" as was requested by the end user. The limit is only active when large changes are required for instance in a start-up as is shown in figure 8. During normal operation, there are no limits in control action.

7. Summary and Discussions

In this paper we have attempted to form a standard for "Edge computing" i.e. how advanced functions created in research tools like Modelica and Python can be included in automation. Such a standard is necessary for broader use. As far as we know, the attempts that have been done so far consist of special programming in each individual case which is much more costly and requires help from a software specialist. With a standard setup as is described here, the process engineer can use external functions made by researchers in the automation system without the assistance of software experts.

The proposed solution showed that it was feasible to implement a MPC strategy together with an existing industrial control system without interfering with its critical functionality and real-time behavior. The execution time for the MPC algorithm was reasonable and it performed well for the specific process. The Modelica simulator was not used in the experimental setup in this case but may be used for more complicated physical-based models as a local "Digital Twin". We have shown that the capacity is sufficient for such solutions.

In this experiment the dynamics of the process being controlled were slow which meant that the CPU capacity was enough to calculate the setpoint once per hour. Processes with faster dynamics may require other approaches, for example other hardware solutions, such as dedicated hardware, a more simplified model for the MPC or different solvers and optimizers. The Python interface to the ICS could also be improved regarding data conversion, messages, optimization etc.

The main idea about using python as a base for MPC, and other types of control algorithms is that implementations easily can be transferred from construction and design phase to an ICS without any further or very little modifications. Python is also a well-known language which has many libraries

available for different purposes and also has a huge amount of support among the developer community. Many other use-cases exist that can take advantages of the introduction of Python in an ICS, for example different ML-implementations which will directly appear as new blocks in the ICS. Using ML in the ICS with data-driven models could be an alternative to the physics-based models implemented in the simulation task. Future research should focus on further validating the framework's performance in different industrial processes and explore the scalability and reliability of the proposed solutions.

Acknowledgment
This work is supported by the District Heat project. Further Thanks to the team at Mimer for their help and support.

References

[1] Schwenzer M, Ay M, Bergs T, Abel D (2021) 'Review on predictive control: an engineering perspective', *The International Journal of Advanced Manufacturing Technology (*2021) 117:1327–1349
.doi:10.1007/s00170-021-07682-3

[2] Krupa P, Limon D,Alamo T (2021) 'Implementation of Model Predictive Control in Programmable Logic Controllers´, *IEEE Transations on control systems technology*, vol. 29, no. 3, May 2021 pp 1117–1130.
doi:10.1109/TCST.2020.2992959

[3] de Oliveira DB, de Oliveira RS (2016) 'Timing analysis of the PREEMPT RT Linux kernel'.*Software-practice & experience* , vol.46 issuse 6 pp 789-819.
doi:10.1002/spe.2333

[4] Adam, GK, Petrellis, N, Doulos, LT (2021). 'Performance Assessment of Linux Kernels with PREEMPT_RT on ARM-Based Embedded Devices´. *Electronics 2021*, 10, 1331.
doi:10.3390/electronics10111331

[5] Carvalho AA, Machado CLD,, Moraes FS (2019). 'Raspberry Pi performance analysis in real-time applications with the RT-Preempt patch´. *2019 Latin american robotics symposium, 2019 Brazilian symposium on robotics (SBR) and 2019 workshop on robotics in education*. pp 162-167
doi:10.3390/electronics10111331

[6] Saletti C, Zimmerman N, Morini M, Kyprianidis K, Gambarotta A (2021). 'Enabling smart control by optimally managing the State of Charge of district heating networks´. *Applied Energy, 283*, p.116286.
doi:10.1016/j.apenergy.2020.116286

[7] Saletti C, Zimmerman N, Morini M, Kyprianidis K, Gambarotta A (2022). 'A control-oriented scalable model for demand side management in district heating aggregated communities'. *Applied Thermal Engineering, 201*, p.117681
doi:10.1016/j.applthermaleng.2021.117681

[8] Zimmerman N, Kyprianidis K, Lindberg CF (2019). 'Achieving lower district heating network temperatures using feed-forward MPC.'. *Materials, 12(15*), p.2465.
doi:10.3390/ma12152465

[9] Monghasemi N, Vouros S, Kyprianidis K, Vadiee A (2022) 'A non-linear gray-box model of buildings connected to district heating systems'. *Energy Proceedings* ,Vol 29
doi.org:10.46855/energy-proceedings-10497

Additional information

[10] First Control (2023) 'XC05 operation manual for researchers and engineers'

Development of a MATLAB-based code for quantification of effective void space in porous pavement

Rebecca Allen[a,*], Berthe Dongmo-Engeland[a], Saja Al-Bata[a]

[a]*Department of Built Environment, Oslo Metropolitan University, Norway*
[*]rebeccaa@oslomet.no

Abstract

Porous pavement is a well-documented, low-impact stormwater management technique. When it comes to design of the top layer, the amount of void space (porosity) is often of interest as it influences both infiltration and strength of the pavement. Laboratory equipment can be used to measure the porosity of core samples, but when more detail is required, other equipment or methods must be used. One such method is to scan the entire sample using a computer tomography (CT) machine and then perform some image processing techniques on the scanned data to reconstruct the sample digitally. While the workflow of scanning and processing to produce the 3D digital twin of porous pavement is not new and can be in fact done by open-source or commercial software, there are still some parts of the process that deserve a deeper investigation, for example binarization and segmentation algorithms applied to the solid-and-void space and void space, respectively. This is difficult to do with commercial software which operates like a black-box, and there needs to be more open-source codes that are user-friendly, extendable, and competitive to what commercial software can do. This work presents a MATLAB-based code that allows for a deeper investigation of how one can accurately and efficiently quantify the effective (or connected) void space of a porous pavement sample from a 3D digital model. We demonstrate the effect of dataset coarsening, which can be used to reduce the computational intensity of the algorithm while preserving accuracy. The code is publicly available online to allow for reproducible research and the possibility of extensions for increased functionality and complexity.

1. Introduction

Asphalt is ubiquitous in city parking lots, roads and highways, and research is often focused on how to increase its lifespan to minimize maintenance costs. However, one problem that arises with asphalt covered surfaces is storm water management: excessive rainfall leads to high volumes of runoff that must be managed properly to avoid surface flooding. Traditional methods to handle this surface flooding involve collecting runoff in storm sewers and moving this water either directly to a body of water (lake or sea) or to an area where it can collect before eventually draining into the ground.

An alternative approach to stormwater management involves the use of asphalt which is typically termed *porous* asphalt (PA) or also *drainage* asphalt (DA) (Dylla & Hansen, 2015), and is typically recommended for parking areas and low-volume roadways (Roseen et al., 2012). As these names imply, the role of this alternative asphalt is to allow surface water to permeate through the layer, from top to bottom, while at the same time remaining strong enough to sustain typical vehicle loads. The challenge is thus the trade-off between a higher-than-typical porosity and reduced strength. When installed as an overlay above traditional impervious asphalt pavements, PA is known as Permeable (or Porous) Friction Course (PFC), Porous European Mix (PEM) or Open-Graded Friction Course (OGFC), a type of pavement developed to improve road safety under wet conditions and reduce noise (Stanard et al., 2007; Watson et al., 2018). Noise reduction and improved safety are the reason why the surface layer of the majority (>90%) of the Dutch principal motorway network consists of a course PA (Aalst et al., 2015; Bondt et al., 2016; Plug & Bondt, 2021).

Advances in PA research can be realized through studying the unique relationship between porosity and strength. Studies (Chen et al., 2021, Ferreira et al., 2021, Król et al., 2017) have focused on obtaining empirically-derived expressions for permeability as a function of porosity. The value of these expressions lies in the fact that they can be used to predict permeability given a certain change in porosity caused by clogging, for example. But first, to capture the PA's ability to filter water, the

connected[1] pore space needs to be determined, i.e., the amount of the total pore space that becomes filled with water as it flows from top to bottom of the layer. The other pore space can be classified as *isolated* since they are not directly connected to the connected pores. Determining the total volume of the connected pore space can be measured in a lab with the right equipment, however information about the exact geometry of the connected pore space is required in order to study the infiltrating- or permeable-behavior of the PA. A popular and non-destructive method to get highly-resolved pore space details involves taking hundreds of 2D images of a PA sample using a computer-tomography (CT) scanner and then performing a set of image processing techniques on these 2D images to construct a 3D digital representation of the sample (Schuck et al. 2021). This same workflow of imaging, processing, and 3D reconstruction of a so-called *digital twin* has also been applied to many other areas of application, such as pervious concrete (Jagadeesh et al., 2022), geomaterials (Quinteros & Carraro, 2023), and even green roofs (De-Ville et al., 2017).

The studies cited above involve independent efforts at preforming image processing techniques in order to extract meaningful information from the CT-scan images. Some use commercial software (Simpleware ScanIP, Avizo) while others develop their own programs or codes that rely on either open-source programming languages and their image-processing tools (Python-based Spam) or license-based programming languages and their image processing tools (MATLAB-based Image Processing Toolbox). While commercial software certainly has its strengths – it is validated, robust, developed by a team of people, and hopefully user-friendly – the efforts of developing a computer program or code to perform research on PA is also valuable and arguably more flexible and *open* than its commercial software competitors. And yet, even though we have found several studies that are based on self-generated code, we find a lack of reproducibility in the research community when it comes to image segmentation applied on PA samples and algorithms used to extract the effective pore space. For example, Schuck et al., 2021 and Jagadeesh et al., 2022 documented very clearly their image processing steps, however leave the reader without any code that could be applied to their work. We understand that it is not always the objective of a researcher to openly publish their developed programs and codes, however we see the benefit that code-sharing can have in the research community. As such, a main objective of this work is to start an online repository (find link at end of paper), for porous media characterization beginning with asphalt samples. Specifically, this study focuses on processing a stack of micro-CT images of asphalt using built-in MATLAB functions and on the characterization of the sample according to its local and global porosity. The code is applied on images from two sources: a laboratory made PA sample and a cored sample from a county road in Norway. The cored sample is not a PA sample but rather considered as regular asphalt and is used to demonstrate the code's ability to quantity porosity in samples with very low porosity.

2. Materials and Methodology

2.1. Asphalt Samples

The asphalt samples studied in this work are shown in Figure 1. "DA" stands for drainage asphalt[2], and "H" stands for Hafjell, a location in Norway where the sample was extracted. The photo for sample 1H is in fact a photo of sample 3H (both of which are studied in Kassem et al., 2023), however it is presented here as a representative photo since they were both cut from the same road. Material details of the DA and 1H samples are summarized in Table 1. The binder type was regular (PGB) for the DA sample and modified with polymer (PMB) for the 1H sample. Dmax stands for the maximum aggregate size.

Figure 1: The porous (left) and regular asphalt (right) samples considered in this work.

Table 1: Material information of samples.

Spec. \ Sample	DA 11 A	1H
Type	Porous	Regular
Source	Lab-made	Cored
Binder type	PGB 70/100	PMB 65/105-60
Dmax	11	16

[1] Some researchers use the term *connected* while some use the term *effective*.

[2] PA (porous asphalt) and DA (drainage asphalt) are used interchangeably in this work.

2.2. Computer Tomography (CT) Scanning

Scanning of the samples was done at the Norwegian Geotechnical Institute (NGI) in Oslo, using a Nikon Metrology XT H-225LC device. Details of the scans are summarized in Table 2. Two-thousand projections were used for the DA sample which means a total of 2000 .tiff images were created and each image contained 2000 by 2000 pixels. However, the sample itself and its volume of interest (VOI) take up only a portion of the images; the number of pixels in the 2D region of interest (ROI) and the number of images that make up the VOI are reported in Table 2, and the total number of voxels is the product of these two. As such, the total number of cells in the VOI is less than $(2000)^3$.

Table 2: CT image specifications and dataset sizes in terms of ROI (region of interest) and VOI (volume of interest).

Spec. / Sample	DA	1H
Voxel size (µm)	62.6	56.3
# of projections	2000	2500
# pixels in ROI	1.87×10^6	2.25×10^6
# images in VOI	901	701
# voxels in VOI	1.68×10^9	1.57×10^9

2.3. Image Processing

The CT scanning procedure produced hundreds of .tiff files for each sample, with a total size of 7.5 and 6 Gigabytes for the DA sample and the regular sample, respectively. These image files were processed using MATLAB release version R2021a, MATLAB's Image Processing Toolbox, and a laptop with Intel(R) i7-1165G7@2.8Hz CPU and 32 GB memory. The basic steps taken to process these image files with the objective of detecting the connected pore space is summarized below:

1. Read each .tiff 2D image file
2. Define a ROI, applied to each image
3. Crop and mask the image data to the ROI
4. Filter out noise from each image with either a median filter and/or a Gaussian filter
5. Detect total pore space in each 2D image using a binarization method, based on either automatic or manual thresholding
6. Construct the 3D volume of total pore space by assembling stack of the 2D total pore spaces
7. Apply the *detect-connected-pores* algorithm on total pore volume to detect the top-to-bottom connected pore volume

The region of interest (ROI) was manually cropped with a draw circle tool such that all the space beyond the sample boundaries, in addition to a small outer ring of sample, was labelled as outside the ROI. This ensured we were focusing only on the sample. The noise reduction filters used were a median filter followed by a Gaussian (with standard deviation equal to 2) for the PA sample and Gaussian (with standard deviation equal to 2) followed by adaptive histogram equalization for the 1H sample. The binarization method was based on manual thresholding for both samples. The output of steps 1 to 5 are presented in Figure 2, and output from steps 6 to 7 in Figure 3.

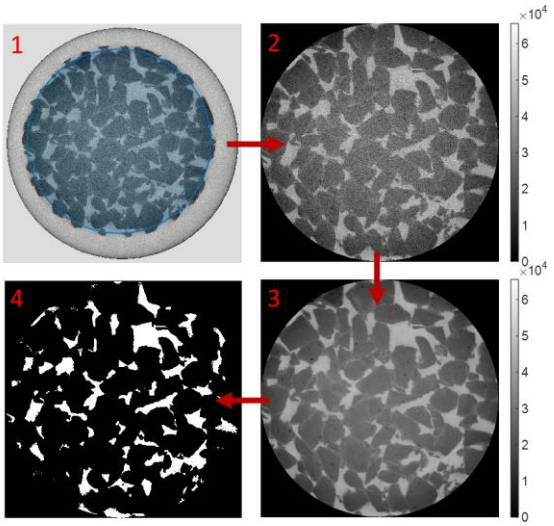


Figure 2: Visualization of the various stages of image processing, from 1-original, 2-cropped and masked, 3-filtered, and 4-binarized image.

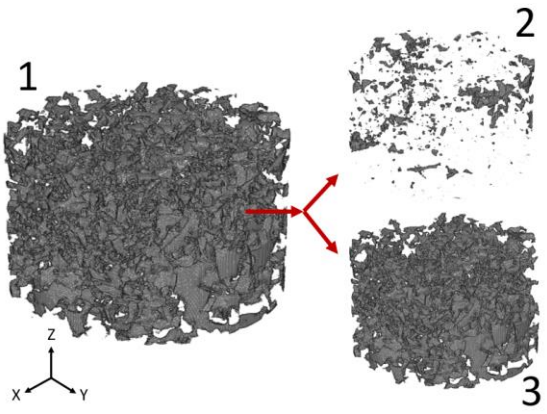


Figure 3: Visualization of a constructed 3D pore volume (1) which is made up of isolated pores (2) and connected pores (3).

2.4. Detecting Connected Pores

The main contribution of this work is the detect-connected-pores algorithm which is openly available online (find link at end of paper). The basic idea of the algorithm is to initialize and continually update a pore cell map, which includes the index of all the pore cells as well as the indices of the 6 neighboring cells. Besides the index, a label is assigned to each pore cells (2=pore needs checking, 1=connected pore). All pore cells in the

top image are immediately labelled with as a top-connected pore. Then a while-loop is used to continually update the labelling of the pore cells based on whether or not they have a neighboring cell that is labelled as top-connected. To help with performance, the neighboring cells that were just used to label a pore cell as top-connected is labelled with NaN so they are skipped over during any consecutive iterations. Two while-loops are required: the first one detects all top-connected pore cells, and the second one detects all bottom-connected pore cells. Two while-loops are used because a pore cell could be connected to other pore cells that eventually lead up to the top of the sample. However, this does not mean they will lead down to the bottom of the sample. Thus, the top-to-bottom connected pores are pores which are common to both the top-connected AND bottom-connected pathways.

3. Results

We present total pore space results for each sample (DA and 1H), since the 1H sample provides a nice comparison to a regular and road-cut asphalt sample. However, afterwards we focus our attention on the DA sample only since the 1H sample lacks a "top-to-bottom" connected pore space.

The connected pore space that was extracted using the detect-connected-pores algorithm on the DA sample is illustrated on the left side of Figure 4. The right side of Figure 4 illustrates all the pore space of sample 1H because a top-to-bottom connected pore space could not be detected in this sample. This is not surprising given that it was not designed to infiltrate water.

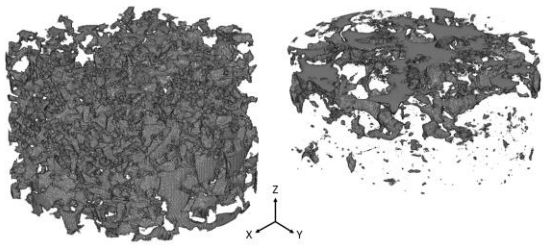


Figure 4: Connected pore space in our DA sample (left) and total pore space in our 1H sample (right). Both samples are fully-resolved (not digitally coarsened).

The vertical distribution of porosity of each sample is plotted in Figure 5. For the DA sample, the top is located at 6 cm and the bottom is located at 0 cm, and a few millimeters from both ends were not included in the digital analysis to avoid boundary effects. The connected porosity is always lower than the total porosity, as expected. While there is some fluctuation in porosity with depth, we notice a slight trend of increasing porosity with sample depth. A possible reason for this behavior is the sample was compacted on only one end, namely the top-end. Thus, the porosity is smaller at the top than at the bottom. Furthermore, the typically observed bathtub shape in the porosity-distribution is not very evident in our DA sample, although it is possible that it could appear in the profile if a few more slices were included on the ends. For the 1H sample, the top is located at approximately 4.3 cm and the bottom is located at 0 cm, and a few millimeters from both ends were not included in the analysis, similar to the DA sample. The vertical porosity distribution in the 1H sample is quite different from the DA sample; it is quite non-uniform and contains a sudden reduction of porosity at about 2.3 cm. That is, the top half of the sample has an average porosity of 0.4 while the bottom half has an average of 0.003. This bottom half contains very few pores, and any pore that does exist is very isolated from its surrounding pores. It is unlikely that water would easily filter through this road's top layer, and of course, it was never designed too.

For comparison, the porosity of the DA sample was measured in the lab using the CoreLock apparatus and found to be 16.6%.

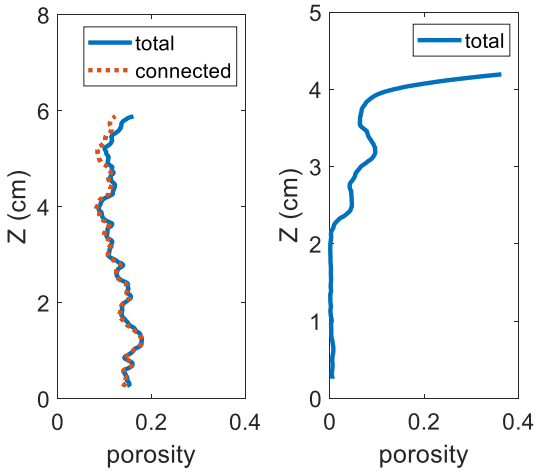


Figure 5: Vertical distribution of porosity in our DA sample (left) and 1H sample (right). Both samples are fully-resolved (not digitally coarsened).

Besides the porosity distributions, we are interested in knowing how fast our detect-connected-pores algorithm is. Our algorithm that detects the top-to-bottom connected pore space is computationally intensive because it uses two different while-loops that must iterate many times until the search for connected cells is over. The first row of Table 3 shows the CPU run time and number of while-loop iterations required in the finest resolution of our DA sample which contained around 218 million pore voxels.

Table 3: CPU run time and while-loop iterations required to detect connected pores in our DA sample in the non-coarsened case (1) and two coarsened cases (15 and 30).

Coarsening level	# pore voxels ($x10^6$)	#connected pore voxels ($x10^6$)	CPU run time (min)	# while-loop its. needed
1	218.05	207.16	1022	5468
15	14.82	13.36	35.7	3360
30	7.56	6.00	17.7	3058

We consider the CPU run time of 1022 minutes problematic. To reduce the amount of time it takes for the detect-connected-pores algorithm to finish, we ran several data coarsening experiments. That is, we coarsened the vertical resolution of the samples, and studied how CPU run time, number of while-loop iterations, total and connected porosity changes with vertical coarsening. Results from the R=15 and R=30 coarsening experiments performed on the DA sample are shown in the rest of Table 3 and the iteration count results from all of the coarsening experiments performed on this sample are plotted in Figure 6.

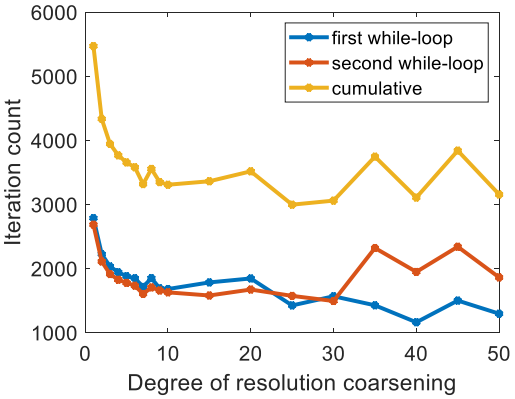


Figure 6: Influence of coarsening on the number of iterations required for the detect-connected-pores algorithm to finish.

Coarsening most certainly reduces the CPU run times, mainly because less while-loop iterations are required before all neighboring-cells to pore-cells have been labelled as NaN (which means they do not get checked again). However, our intention is to maintain sufficient accuracy of the digitally measured porosity values, and thus realize we cannot over-coarsen the datasets. The vertical distributions of porosity under R=1, R=15, and R=30 are shown in Figure 7. These profiles show that coarsening smooths out some of the perturbations in both the total and connected profiles, and that the connected profile moves further to the left of the total profile as the degree of coarsening increases. This means the pore space is less connected as the sample dataset is coarsened. Going back to Figure 6, there is a sudden change in iteration count behavior at around R=30 and above: it is likely that over-coarsening has resulted in a connected pore structure that looks quite different from the previous coarsening degree. Skipping over the key pores that link the majority of the connected pore space together can produce a very different result.

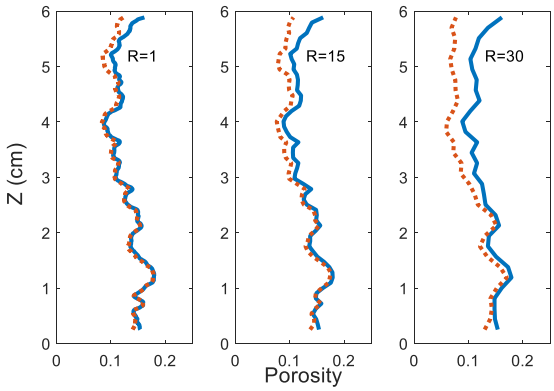


Figure 7: Total (solid line) and connected (dashed line) porosity distribution in the DA sample dataset that was coarsened by various degrees: R = 1, 15, 30.

To better quantify the effect of coarsening, we compare the porosity values in each coarsened sample (R=2 to 50) against the porosity values in the finest-resolved sample (R=1). That is, we calculate the relative difference between coarsened to non-coarsened (or finest) results. Results for the DA sample are presented in Figure 8.

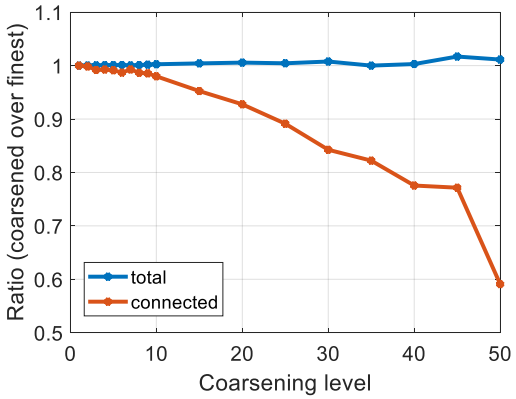


Figure 8: Impact of resolution coarsening on the digitally measured total and connected porosity in the DA sample. Results are given in terms of porosity divided by finest porosity.

This plot indicates the following three things. First, coarsening tends to overestimate total porosity and at the same time underestimate connected porosity. Second, accuracy is not greatly sacrificed under coarsening of R=2 to 10 (up to 0.5% relative difference for total porosity and 2% relative difference for connected porosity; see R=10). But applying coarsening beyond R=10 results in up to

1% relative difference for total porosity and 40% relative difference for connected porosity; see R=50. And third, coarsening impacts the connected porosity measurement more than the total presumably because more pores have become isolated from the top-to-bottom pore pathways. The pores along the perimeter of the cylindrical sample could be most suspectable to isolation.

4. Summary

This work presents an algorithm to detect the connected pore space of porous asphalt. This algorithm was applied on a real porous asphalt sample and the pore space was characterized in terms of its 2D porosity distribution in the vertical direction. The algorithm was assessed in terms of the time it took to analyze the dataset comprised of about 200 million pore cells. To reduce the computational intensity of analyzing 200 million cells, the dataset of the DA sample was vertically coarsened and acceptable estimates of the connected pore space were obtained in a much shorter length of time. Specifically, a coarsening level of 10 cut the CPU time by a factor of around 25 while producing a digital representation that was only 2% difference relative to the full resolution result. The detect-connected-pores code is available online; see link presented in the Author Notes section below. Future work is planned to develop the repository further to include code used to simulate water filtration through the pore space and characterize fluid flow parameters such as permeability and tortuosity, as well as to calculate parameters that include mechanical characteristics of the asphalt such aggregate-to-aggregate contact, binder layer thickness, etc.

Author Notes

The online repository we created for porous asphalt analysis and used in this work is available at: https://github.com/AllenGitCode/PorousAsphalt-ImageProcessing.

Acknowledgment

We thank Statens Vegvesen (The Norwegian Public Roads Administration) for making the asphalt samples used in this work, as well as for performing the laboratory tests on the samples.

We thank NGI (Norwegian Geotechnical Institute) for the use of their CT scanner to obtain the image datasets.

Part of this work was funded by Konnekt (National competence center for transport in Norway) under project code/title: 202640/Collaborative project on transport infrastructure.

References

Aalst, W. van, Derksen, G., Schackmann, P.-P., Paffen, P., Bouman, F., & Ooijen, W. van. (2015). Automated Ravelling Inspection and Maintenance Planning on Porous Asphalt in the Netherlands. International Symposium Non-Destructive Testing in Civil Engineering (NDTCE 2015). Berlin.

Bondt, A., Plug, K., van de Water, J., The, P., & Voskuilen, J. (2016). Development of a durable third generation Porous Asphalt with a high noise reduction. doi: 10.14311/EE.2016.171

Chen, S., You, Z., Yang, S.-L., Garcis, A., Rose, L. (2021) Influence of air void structures on the coefficient of permeability of asphalt mixtures. Powder Technology, 337, 1-9. doi: 10.1016/j.powtec.2020.08.082

De-Ville, S., Menon, M., Jia, X., Reed, G., Stovin, V. (2017) The impact of green roof ageing on substrate characteristics and hydrological performance. Journal of Hydrology, 547, pp. 332-344. doi: 10.1016/j.jhydrol.2017.02.006

Dylla, H., & Hansen, K. R. (2015). Porous asphalt pavements with stone reservoirs.

Ferreira, W.L.G., Branco, T.F.C., Caro, S., Vasconceios, K. (2021) Analysis of water flow in an asphalt pavement surface layer with different thicknesses and different permeability coefficients. Road Materials and Pavement Design, 22(1), 82-100. doi: 10.1080/14680629.2019.1617186

Jagadeesh, A., Ong, G. P., & Su, Y. M. (2022). Selection of image processing algorithms for evaluation of pervious pavement pore network properties. In Road and Airfield Pavement Technology: Proceedings of 12th International Conference on Road and Airfield Pavement Technology, 2021 (pp. 559-571). Cham: Springer International Publishing.

Kassem, A., Linnerud, A.H., Martinussen, T.S. (2023). Digitale tvillinger for kvalitetskontroll av asfaltdekker. Bachelor's thesis, Oslo Metropolitan University.

Król, J. B., Khan, R., Collop, A. C., (2017) The study of the effect of internal structure on permeability of porous asphalt. Road Materials and Pavement Design. doi: 10.1080/14680629.2017.1283355

Plug, K., & Bondt, A. (2021). Maintenance and Rehabilitation OPA8-Inlay as heavy maintenance measure for OPA8 Porous Asphalt.

Roseen, R. M., Ballestero, T. P., Houle, J. J., Briggs, J. F., & Houle, K. M. (2012). Water quality and hydrologic performance of a porous asphalt pavement as a storm-water treatment strategy in a cold climate. Journal of Environmental Engineering, 138(1), 81–89.

Quinteros, V.S., Carraro, J.A.H. (2023) The initial fabric of undisturbed and reconstituted fluvial sand. Geotechnique, 73(1), (pp. 1-15). https://doi.org/10.1680/jgeot.20.P.121

Stanard, C., Candaele, R., Charbeneau, R. J., & Barrett, M. E. (2007). State of the Practice Permeable Friction Courses.

Schuck, B.,Teutsch, T., Alber, S., Ressel, W., Steeb, H., & Ruf, M., (2021) Study of air void topology of asphalt with focus on

air void constrictions – a review and research approach. Road Materials and Pavement Design, 22:sup1, (pp. S425-S443). https://doi.org/10.1080/14680629.2021.1907215

Watson, D., Tran, N. H., Rodezno, C., Taylor, A. J., James, T. M., National Cooperative Highway Research Program, Transportation Research Board, & National Academies of Sciences, Engineering, and Medicine. (2018). Performance-Based Mix Design for Porous Friction Courses (p. 25173). Transportation Research Board. https://doi.org/10.17226/25173

Machine learning assisted real-time heat load forecasting in district heating network

Avinash Renuke [a,*], Stavros Vouros [a], and Konstantinos Kyprianidis [a]

[a] *Mälardalen University, Västerås, Sweden*
*corresponding.avinash.renuke@mdu.se

Abstract

District heating system often consists of a long, complex network of piping carrying heat from a power plant to the consumers. The supply temperature from the plant is either controlled by the operator from experience or a predefined curve based on the outdoor temperature. An optimized supply temperature which would be lower than the one obtained traditionally would lead to lower heat loss and reduced peak load on the power plant. In this paper, we investigate the machine learning models for heat load forecasting which is a crucial parameter in the optimizing process. Models are generated using supervised machine learning algorithms: Linear models (Linear Regression, Ridge and Gaussian Process Regressor), Random Forest Regressor, Support Vector Machine (SVM) and Long Short-Term Memory (LSTM) recurrent neural network (RNN). Data-driven models are used extensively in the literature to predict heat load prediction based on the weather and the time effect on a fixed training set, however, in this study, we model the heat load in the network in real-time scenarios i.e., adaptive training and forecasting. The model is adaptively updated as well as the training of the machine learning model in real time. It provides a "plug-and-play" solution for real-time prediction without significant pre-tuning requirements. The results of all the models are compared with various time horizons i.e., 6 hrs, 10 hrs, 24 hrs and 1 week, using the district heating data obtained for the city of Vasteras in Sweden. The performance of the prediction algorithms is evaluated using Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE). An algorithm with the best accuracy is selected based on the performance comparison. Also, models suitable for short-term and long-term forecasting are discussed towards the end of the article.

1. Introduction

Optimizing district heating systems using machine learning (ML) techniques has gained significant interest in recent research. ML models can help improve the efficiency and operation of district heating systems by optimizing heat distribution and consumption. ML models can be used to predict heat demand patterns in buildings or areas connected to the district heating system. By analyzing historical data, weather conditions, and other relevant factors, these models can forecast heat demand accurately. This enables more efficient planning and optimization of heat supply from the plant and distribution, reducing energy waste and costs.

The focus of this article is a combined heat and power (CHP) plant, also known as a cogeneration plant, a type of district heating (DH) system that simultaneously generates electricity and useful heat from a single energy source. The CHP plant uses a primary energy source, such as natural gas, biomass, or waste heat, to produce electricity and heat in a combined process. The primary energy source drives a generator to produce electricity, while the waste heat generated during electricity generation is captured and used for heating purposes. The waste heat produced during electricity generation in the CHP plant is captured and utilized for district heating. This waste heat is typically recovered through heat exchangers and transferred to a heat distribution network. The heat distribution network consists of insulated pipes that transport hot water from the CHP plant to connected buildings and areas. Heat exchanges and control valves regulate the flow and temperature of the heat within the network. Buildings and residential areas connect to the district heating network through heat exchangers in the consumer substation. Heat exchangers transfer the heat from the hot water in the network to the building's internal heating system, providing space heating and sometimes domestic hot water. Currently, the supply temperature from the plant is governed by the operator's expertise and knowledge based on historical consumption data. Due to temperature delay in the piping network, the operator injects heat into the network without any clarity on actual heat demand at the end user. The drawback of this

strategy is that, although satisfying consumer expectations, the return temperature in the network would be higher than necessary. This suggests that lowering the network supply temperature has a significant potential to reduce the load on DH plants in addition to the fact that the supply temperature is frequently greater than is necessary. The optimum peak supply temperature, which is lower than the historical supply temperature, would facilitate more electricity production for the CHP plant. Such energy optimization could result in a district-wide reduction in greenhouse gas emissions given the rising worries about climate change.

Heating load forecasting plays an essential role in reducing network losses and performance optimization (reduced plant supply and return temperature). Supervised machine learning (SL) techniques are widely researched for heat demand forecasting in district heating systems while Reinforcement learning(RL) is suitable for optimal control and load balancing strategies (Idowu, Åhlund and Schelén, 2014). These methods leverage historical data, weather information, and other relevant factors to predict future heat demand accurately. Researchers have studied various ML models such as Regression models, Support vector Machine (SVM), Nature inspired, Artificial Neural network (ANN) etc. (Idowu *et al.*, 2014) presented a data-driven heat load forecasting for multi-family buildings using four ML methods – SVM, ANN, Multiple Linear regression(MLR) and Classification and Regression Tree (CART). The forecasting model was evaluated for horizon values of 1, 3, 6, 12, 18 and 24-h. The SVR method was found to be the best performing followed by MLR. The authors were able to achieve a 5.6% (best-performing) Mean Absolute Percentage Error (MAWP). The use of a context-based regression approach model for average and individual user consumption is shown to be effective (Rongali *et al.*, 2015). Dalipi et al (Dalipi, Yildirim Yayilgan and Gebremedhin, 2016) have performed a comparison of ML models for heat load forecasting for multiple buildings and showed SVR to be the best-performing one. Idowu et al (Idowu *et al.*, 2016) evaluated and compared different ML methods such as SVM, FFNN, MLR and regression trees. The models are produced and evaluated using data observed from 10 district heating substations for five multi-family apartments and five commercial buildings. They included outdoor temperature, day of the week, hour of the day, historical values of thermal load and the physical parameters of a substation (supply temperature, difference between supply and return temperature and flow rate) for a forecast horizon up to 48 hr. They found that SVM gave the best prediction performance, with FFNN and MLR having similar error rates as SVM and Regression tree models with higher performance error rates. Suryanarayana et al (Suryanarayana *et al.*, 2018) have shown that the linear models can be very powerful for heat load forecasting outperforming some of the advanced models like SVR (Support Vector Regressor) and Gradient-Boosted Trees (GBT). They achieved an MAWP of as low as 8.77 with linear models and 8.07 with DNN. Ntakolia et al (Ntakolia *et al.*, 2022) did a comprehensive review of machine learning models used for heat load forecasting and concluded that ANN and SVM are found to be the most frequent models used for heat load prediction.

Accurate heating load forecasting is a precondition to the optimization and control of district heating systems. Zhao et al. (Zhao, Li and Shan, 2021) have used SVM in their study to forecast heat load to optimize the DH system using Model predictive control (MPC). They have shown a reduction in energy peak and total energy consumption of the system. Similarly, Zimmerman et al. (Zimmerman, Kyprianidis and Lindberg, 2019) demonstrated a reduction in network supply and return temperatures by using MPC with feed-forward with CHP heat load reduction of 12.5% to 13.7%. However, the weakest point of the paper was that a realistic heat-load model was missing. The machine learning models studied in the literature for heat load predictions are for fixed training sets and not adaptive for real-time scenarios which creates the motivation for the present study.

The objective of this paper is to compare machine learning models to forecast heat load for the DH system of the city of Västerås in Sweden, for adaptive and real-time forecasting. The model created should not require any pre-tuning for use in the DH system. Following machine learning techniques have been considered in this paper:

1. Linear models: linear regression (LR), Ridge Regression(RR), Gaussian Process regression(GPR)
2. Support Vector Regressor (SVR)
3. Random Forest (RF)
4. LSTM (Long-Short-Term Memory)

The model is developed for a single substation (Tillberga) and then validated with another substation (Skultuna) present in Västerås city (explained in the next section).

The paper is organized as follows. Firstly, the DH system is described followed by a modelling approach and an overview of the considered algorithms is presented. Secondly, the methodology is described, including the training, validation and test periods. The accuracy of the different models is then investigated and compared using experimental data from a real-world DH plant.

2. Methodology

In this section, to emphasize the motivation behind this work, a brief overview of the DH system is provided first. The dataset, ML models and methods, variables, and their structure are then explained.

2.1. District heating system

The DH system presented in this study is from the city of Västeås and its surrounding region. A typical DHS has three main parts - The heat generation - which usually consists of a co-generation plant and/or a heat-only boiler station, the Distribution network - which consists of insulated pipes of varying diameters carrying hot water through the entire network and the substations - where heat is transferred from the primary to the secondary network via a heat-exchanger. Fig. 1(top) shows the basic schematic drawing of a CHP plant and a DH system. The red and blue lines denote supply and return pipes respectively. Fig. 1(bottom) shows the schematic illustration of the city of Västerås with a CHP plant connected with different regions around the city.

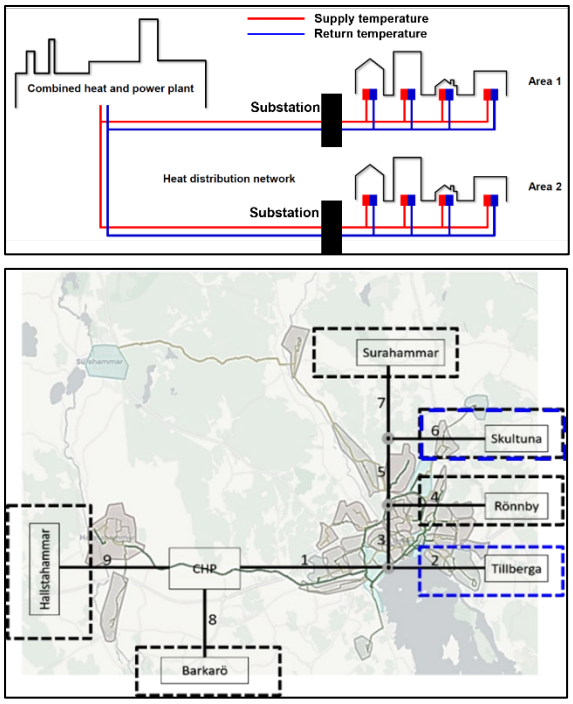


Figure 1: (top)A schematic diagram showing a district heating system network with substations connected to the regions (bottom) Schematic illustration of regions connected with the DH system to the CHP in the city of Västerås and its surrounding regions

The supply temperature of the hot water is controlled directly from the plant's control room based on the outdoor temperature and it follows mostly a given operation temperature curve. The return temperature, on the other hand, depends mainly on the customer's heat usage. The current DH system is 3rd generation where the temperature level varies between 70 and 120°C, particularly during the winter season. The heat load in district heating systems is the sum of all heat loads that are connected to the network and distribution and other losses in the network.

The heat load ML model is created using the DH system of the region Tillberga, which is approximately 14.5 km from the CHP. The reasons for choosing Tillberga are due to the fact of having access to historical network measurements for multiple years. The ML model will be validated both with Tillberga and Skultuna of the city of Västerås.

The heat consumption at the substation, Q_{sb} delivered to the region is mainly a function of the supply temperature, Ts, the return temperature, Tr, and the flow rate, m, as shown in Eq. (1).

$$Q_{sb} = m.C.(T_s - T_r) \quad (1)$$

Where, Q_{sb} is heat consumption at the substation, C is the specific heat of hot water, T_s is supply temperature and T_r is the return temperature at the substation.

2.2. Dataset

The data used in this study are provided by Mälarenergi who operates the CHP plant. These data are measured and collected by regular measurements that are part of the control system in a DH plant. The measurements consist of the time of day, supply temperature (ST), return temperature (RT), flow rate (FR), and heat load (HL) at the individual substations with 1-hour time step intervals. The data used in this study are for the winter season: January 1st 2019 to March 31st 2019 for model selection and January 1st 2022 to March 31st 2022 of Tillberga and Skultuna for testing.

Before building a machine learning model, it is important to preprocess the data and remove or replace any missing values or outliers. The data is preprocessed to remove and replace any missing data/outliers (outside ±3-standard deviation) that could cause problems with the ML model, such as biased results or inaccurate forecasting. We used interpolation to replace the missing/outlier data. Another step in the data preprocessing is to scale the data. Scaling means changing the range of data so that all the values are within a similar range. The data is scaled using the 'MinMaxScaler() (*sklearn.preprocessing.MinMaxScaler*)' function from the Python library which scales and translates each feature individually such that it is in the given range on the training set, e.g. between zero and one.

The data is then split into training, validation, and test sets. This is because we want to train our models on the training data, validate them on the validation data, and then test them on the test data.

In this model, we have used a 70-30 split for training and testing respectively to assess the model.

2.3. Input variables

The heating demand of a consumer depends on several types of factors such as climate and weather conditions, timestamp information, and historical information. In this work, we use the data which is readily available from most of the plants which are limited to weather, historical, and time-stamp information.

a. Historical heat load data
b. Weather input variables: Outdoor air temperature
c. Time-stamp variables: hour of the day and day of the week

2.4. Machine learning models

The following ML models are studied in this paper which are found to be more suitable for the heat load forecasting problem (Ntakolia et al., 2022). In this study, we have used supervised ML models which use a set of input variables to forecast the values of an output variable. In these models, we look at historical data to *train* a model to learn the relationships between variables and a *target (output),* the thing we're trying to forecast. This way, when new data comes in, we can use the input values to make a good prediction of the output (target). We have studied the four most widely used supervised learning algorithms: Linear models (LM), Random Forest (RF): a Tree-based model, SVM and LSTM: ANN based model.

Linear models: Linear ML models are simple (old-school) algorithms that make predictions based on linear relationships between the input variables and the target variable. The relationship between the inputs and the target is represented by a linear function. The following three LMs are used in this study:

a. LR: Linear regression models the relationship between the inputs (features) and target by fitting a linear equation to the observed data. It aims to minimize the sum of squared residuals between the predicted and actual target values.
b. Ridge Regression (RR): Ridge regression is a regularized version of LR that adds a penalty term to the cost function, aiming to reduce the model's complexity and prevent overfitting.
c. Gaussian Process Regression (GPR): GPR is a non-parametric Bayesian regression technique that can be used for regression. It models the relationship between the input and the target variable as a distribution over functions rather than a single function.

Random Forest (RF): RF is a tree-based algorithm that combines multiple decision trees to make predictions. It belongs to the ensemble learning family, where multiple models are combined to improve overall performance and generalization.

Support Vector Regressor (SVR): SVM, based on a statistical learning theory, are one of the most successful and widely applied ML methods, for solving regression problems. SVR is a method of SVM for regressions. SVR is a powerful approach for handling non-linear regression problems by mapping the input variables into a higher-dimensional space and finding an optimal hyperplane that fits as many data points as possible within a specified margin.

Long-Short-Term Memory (LSTM): LSTM is a type of recurrent neural network (RNN) architecture that is widely used for sequential data analysis, such as time series forecasting, natural language processing, and speech recognition. They introduce specialized memory cells and gating mechanisms that allow the network to selectively remember or forget information over time. LSTM networks have been proven effective in modelling and predicting sequences with long-term dependencies (Schaefer, Udluft and Zimmermann, 2008).

Hyperparameter analysis is a crucial step in machine learning to optimize model performance. It involves tuning the hyperparameters of a machine learning algorithm to find the best combination that yields the highest accuracy or lowest error on a given dataset. In this study, an automated grid search was conducted for each specific ML model by selecting the hyperparameters of the ML model. Grid search is a popular technique for hyperparameter optimization in machine learning. It involves exhaustively searching through a specified grid of hyperparameter values to find the combination that yields the best model performance.

Table 1 shows the hyperparameters used in each ML model.

All the ML models are assessed using open-source software like Python with packages like Scikit-Learn (*scikit-learn: machine learning in Python — scikit-learn 1.2.2 documentation*, 2023) and Tensorflow (*TensorFlow*, 2023).

2.5. Performance metrics

In this work, we considered commonly used metrics for evaluating the performance of the proposed models. These are the Root Mean-Square

Error (RMSE), the Mean Absolute Percentage Error (MAPE) and the Correlation coefficient (Corr Coef). In this paper, we have used MAPE and RMSE for performance evaluation.
RMSE is commonly used to measure the difference between a model's predicted values and actual values observed (the average prediction error over all time instants) and is computed as:

$$RMSE = \sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2\right)} \quad (2)$$

Table 1: ML models and hyperparameters

ML model	Hyperparamete	Definition	Defined Parameter
LR (Linear	-	-	-
RR (Ridge Regression)	alpha (α)	the regularization parameter	0.01 - 10
GPR (Gaussian Process Regression)	Kernel type	the Kernel function	RBF(Radial Basis Function)
	length_scale	Smoothness of the kernel	0.1- 10
	alpha (α)	Regularization Parameter	0.001-10
RF (Random Forest)	n_estimators	number of decision trees in the random	100-500
	max_depth	the maximum depth of each	0-10
SVR (Support Vector Regressor)	Kernel type	the Kernel function	RBF(Radial Basis Function)
	C	the penalty paramet	0.1- 10
	epsilon	the margin of tolerance around the predicted value	0.01-1
LSTM (Long-Short-Term Memory)	units	the number of memory cells or hidden units in the LSTM layer	10-500
	activation	Activation Function	relu
	Batch Size	-	8-64
	Number of Epochs	-	50-500
	Loss Function	-	mean_squared_error
	optimizer	-	adam

MAPE estimates how close forecast values are to actual values in percentage and is computed as:

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\frac{|\hat{y}_i - y_i|}{|y_i|} \quad (3)$$

where $\hat{y}_i$ is the estimated value in a forecasting model, y_i is the measured value and n is the total number of forecasted data points in each forecast horizon.

2.6. Model workflow

As mentioned earlier, we have used two datasets from different years to build and test the models in adaptive learning and real-time scenarios. The first dataset i.e. dataset from region Tillberga from 2019 is used to test the performance of different ML models. The decision is taken based on their performance metrics and the best model is selected for adaptive learning. Figure 3 shows the workflow of the development of the heat load forecasting model. The selected model with optimized hyperparameters will be used in the dataset of the year 2022 for Tillberga and Skultuna of Västerås City. The final model is trained adaptively as the new data adds to the historical set and the training set. The training set moves as the new data is available i.e., the model which dynamically adapts to the new patterns in the data.

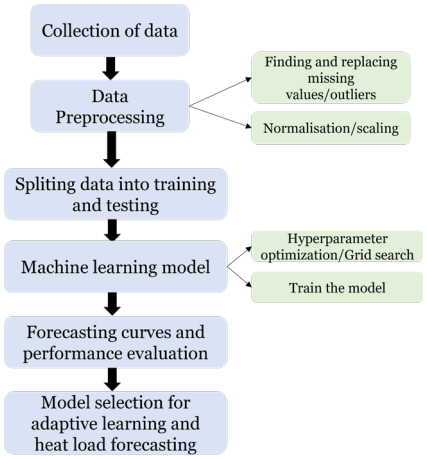


Figure 2: Workflow for machine learning model selection and heat load forecasting

3. Results

In this section, results are presented in two subsections: a) the performance of different models and selection for the testing data b) the performance of selected models in adaptive learning.

3.1 Performance comparison of ML models

The performance of all the ML models mentioned above is evaluated for the dataset of the Tillberga substation. The data is from the winter season from 1 January 2019 to 31 January 2019. Table 2 shows the performance of different ML models for the testing dataset. Amongst the linear models, GPR shows very good performance in prediction with 4.77% MAPE with LR and RR showing similar performance. GPR can capture the complex and non-linear relationship between input features and output. The high accuracy of GPR suggests that there might be a non-linear relationship between features and output.

Table 2: ML models and hyperparameters

ML model	MAPE, %	RMSE
LR (Linear Regression)	8.22	394.8
RR (Ridge Regression)	8.26	388.5
GPR (Gaussian Process Regression)	4.77	242.4
RF (Random Forest)	4.30	233.8
SVR (Support Vector Regressor)	4.28	225.8
LSTM (Long-Short-Term Memory)	5.50	279.9

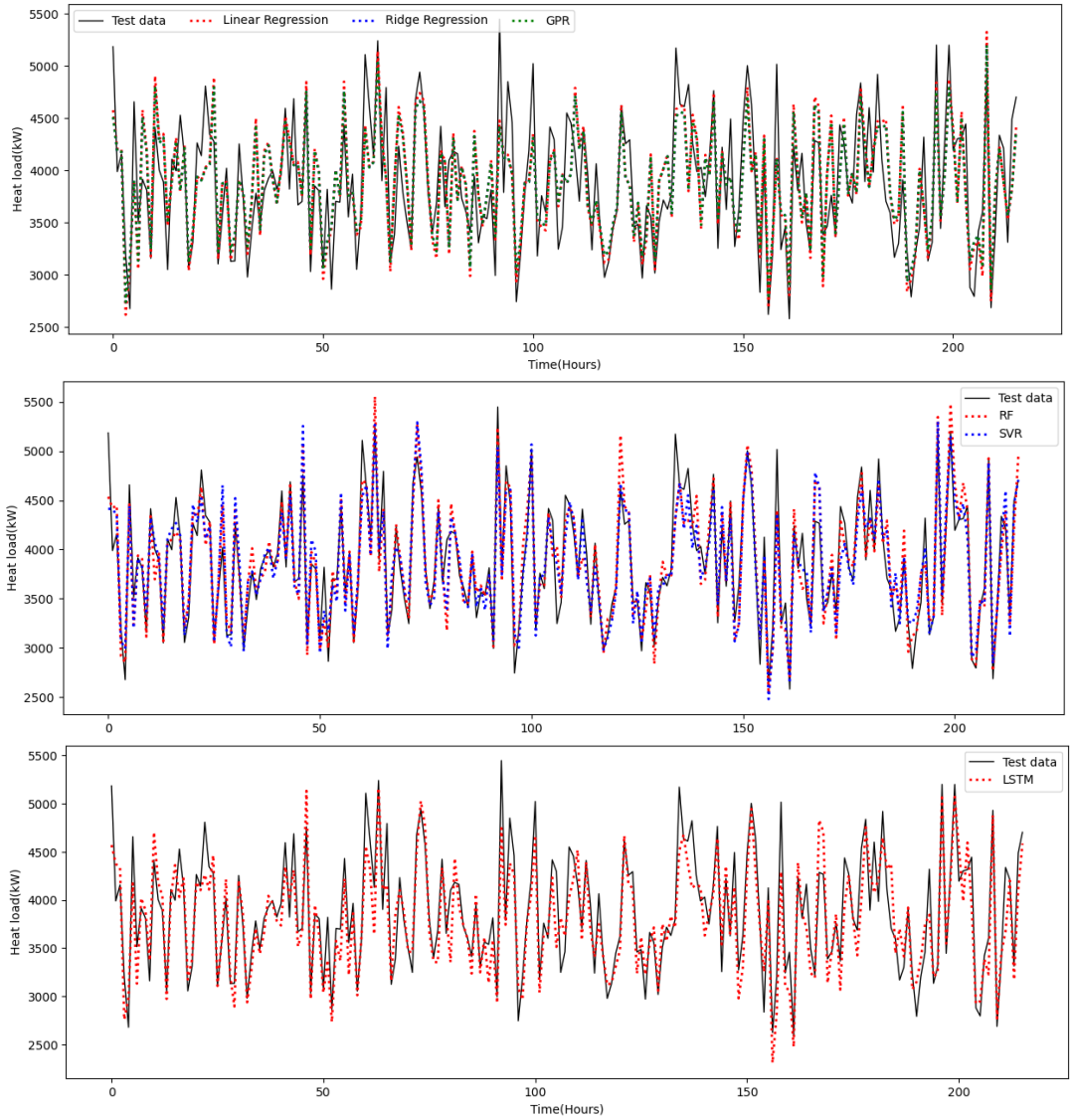


Figure 3: Performance comparison of actual and forecasted heat load for testing data for ML models

Table 1 also shows that the best-performing model is SVR with 4.28% MAPE which suggests that predicted heat loads with SVR are closer to the actual heat consumption as also seen in Figure 3. The best performance of the SVR over the other methods is associated with efficient feature space modelling and the fact that SVR is less prone to overfitting. It can be concluded that the GPR, RF and SVR can be effectively applied to the prediction of heat load in the DH system. In this study, we select SVR for adaptive training in real real-time scenario of heat load forecasting.

3.2 Adaptive modelling and performance

The SVR model is applied to the dataset of Tillberga and Skultuna region of Västerås city for adaptive learning and heat load prediction. The data is taken for the winter season from 1st January to 31st March 2022. The performance is evaluated for the adaptive training set of two weeks and four weeks. Figure 4 shows an example of the performance execution for the Tillberga region. The Fig, black line represents the measured data of the heat load of Tillberga against which predicted values (green line) are plotted. The training set is shown in the red line which adaptively moves with the addition of new data. The blue line represents the predicted horizon (6 hrs, 10 hrs, 24 hrs or 1 Week). The hours in the period are plotted on the x-axis which represents the total time duration.

Figures 5 and 6 show results of actual heat load and heat load prediction for 24 hrs, based on the SVR algorithm for Tillberga and Skultuna respectively. The figures also show the performance indicators (MAPE and RMS) for two different training set lengths i.e., 2-weeks and 4-weeks.

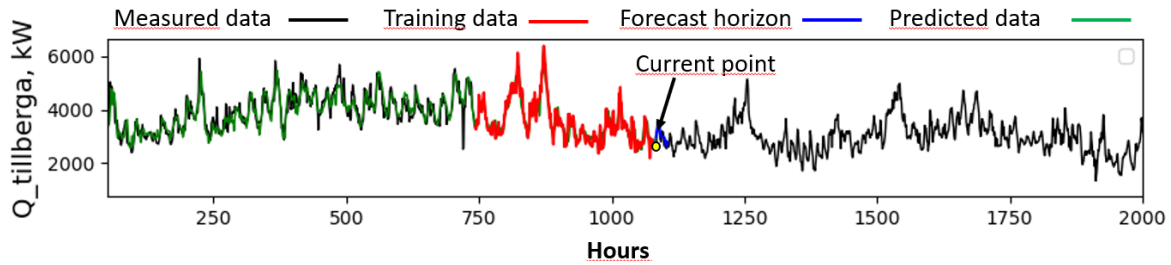


Figure 4: Adaptive learning of model to predict heat load

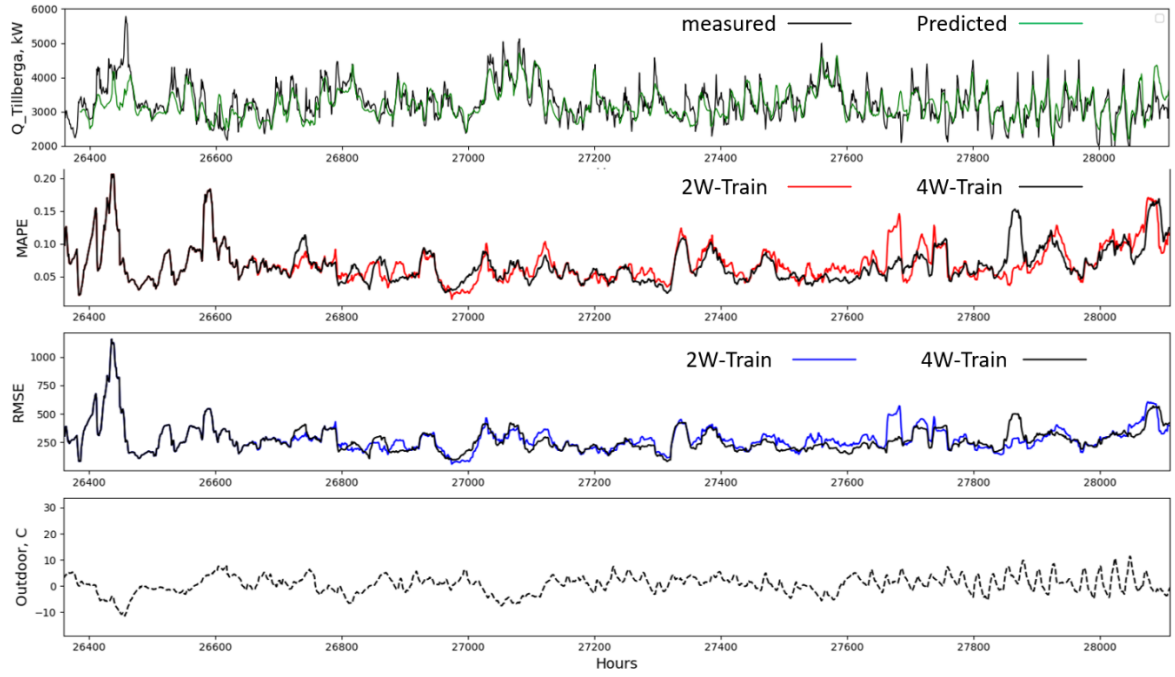


Figure 5: Forecasted heat load and Performance indicators (MAPE and RMSE) for **Tillberga** region with data from 1st Jan 2022 to 31st March 2022 with **24 hr** of prediction horizon

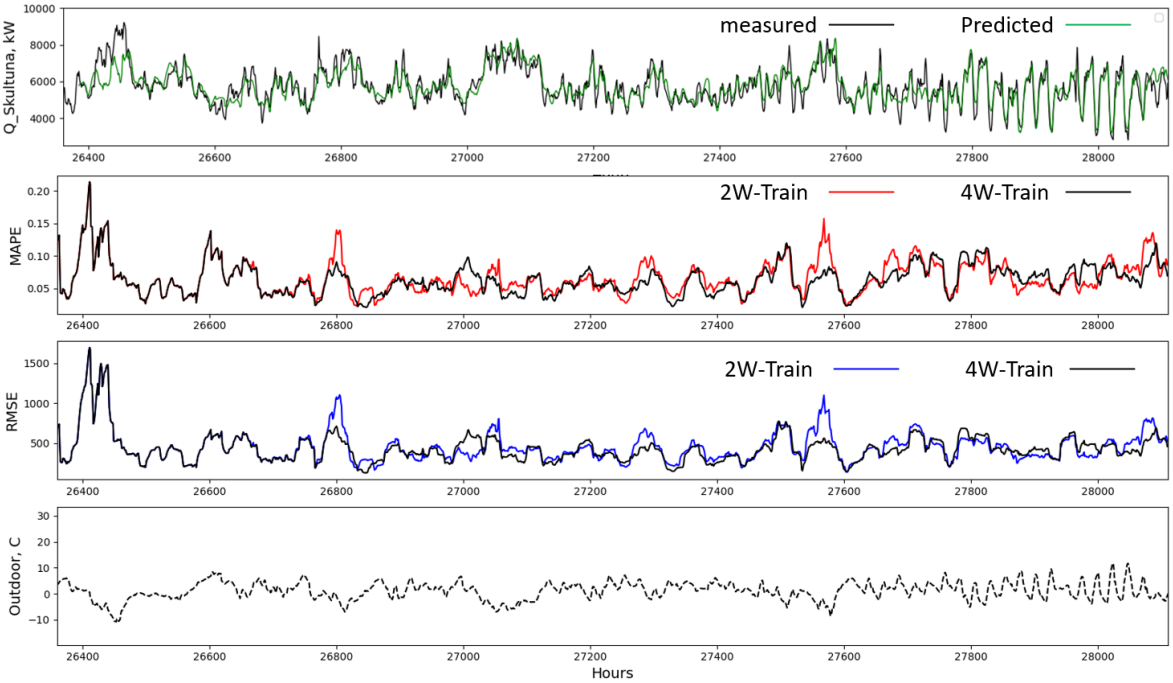


Figure 6: Forecasted heat load and Performance indicators (MAPE and RMSE) for the **Skultuna** region with data from 1st Jan 2022 to 31st March 2022 with **24-hour** prediction horizon

As can be seen from Figure 5, the predicted HL with SVR is closer to the actual energy consumption, with a MAPE value fluctuating between 3 to 10%. The high error at the initial predictions is due to the small training set available for the forecasting model. However, as the training set achieves the 2-week/4-week data, the accuracy improves. Also, the 4-week training data shows better performance than the 2-week training data which follows the intuition. The high error shown by the 2-week training data at around hour 27650 is due to the exclusion of the low outdoor temperature data. The 4-week training data captures the low outdoor temperature hence the more accurate predictions. The RMSE plot shows similar behaviour, and no deviations are observed. This confirms the good performance of the SVR for adaptive learning with good accuracy.

The model is adaptively updated as well as the training of the machine learning model in real time. It provides a "plug-and-play" solution for real-time prediction without significant pre-tuning requirements. This claim is validated by applying the SVR model to a completely different data set of another region i.e., Skultuna. Figure 6 shows the forecasted heat load and Performance indicators (MAPE and RMSE) for data from 1 Jan 2022 to 31 March 2022. The performance is quite like seen in the previous case of Tillberga region. The MAPE is within the range of 3-10%.

The performance discussed above was for 24 24-hour forecast horizon. The performance with 6 hours, 10 hours and 1 week is also evaluated and presented in the appendix section of the paper. It can be seen that for the short forecasting horizons, 6 hrs and 10 hrs, the errors are slightly higher than the long forecast horizon. The error in the long forecasting horizon seems to be more stable.

4. Conclusion

In this paper, six ML algorithms for the heat load prediction in the DH network of Västerås city are developed, compared and analyzed. The algorithms studied are LR, RR, GPR, RF, SVR and LSTM. Hyperparameter analysis is carried out to find the optimized values for each model. The performance of the models was compared using data from 2019 for the winter season. The predicted hourly results were compared with actual heat load data.

The SVR algorithm proved to be the most efficient one, producing the best performance in terms of MAPE and RMSE. The SVR model is then selected for adaptive learning and heat load forecasting in real-time scenarios. The model is tested for the actual data from the winter of 2022 of the Tillberga region. The results are also compared with shorter (2-week) and longer (4-week) training sets. The model, overall, shows good performance with MAPE ranging from 3 to 10%. To provide a "plug-and-play" solution for real-time prediction without significant pre-tuning requirements, the model is tested also with completely different regions on winter data. The performance shows similar accuracies to that in the Tillberga region. It proves that the developed SVR method is appropriate for adaptive learning and application in heat load prediction. In future, we intend to use this model to predict heat load in real-time scenario for the optimization of DH network supply and return temperature.

Acknowledgement

The authors would like to thank Mälarenergi for providing the data used for this work.

This research was performed as part of the "DISTRHEAT—Digital Intelligent and Scalable control for Renewables in HEAting neTworks" project, which received funding in the framework of the joint programming initiative ERA-Net Smart Energy Systems' focus initiative Integrated, Regional Energy Systems, with support from the European Union's Horizon 2020 research and innovation programme under grant agreement No 775970.

References

Dalipi, F., Yildirim Yayilgan, S. and Gebremedhin, A. (2016) 'Data-Driven Machine-Learning Model in District Heating System for Heat Load Prediction: A Comparison Study', *Applied Computational Intelligence and Soft Computing*, 2016, pp. 1–11. Available at: https://doi.org/10.1155/2016/3403150.

Idowu, S. *et al.* (2014) 'Forecasting heat load for smart district heating systems: A machine learning approach', in *2014 IEEE International Conference on Smart Grid Communications (SmartGridComm)*. *2014 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, Venice, Italy: IEEE, pp. 554–559. Available at: https://doi.org/10.1109/SmartGridComm.2014.7007705.

Idowu, S. *et al.* (2016) 'Applied machine learning: Forecasting heat load in district heating system', *Energy and Buildings*, 133, pp. 478–488. Available at: https://doi.org/10.1016/j.enbuild.2016.09.068.

Idowu, S., Åhlund, C. and Schelén, O. (2014) 'Machine learning in district heating system energy optimization', in *2014 IEEE International Conference on Pervasive Computing and Communication Workshops (PERCOM WORKSHOPS)*, pp. 224–227. Available at: https://doi.org/10.1109/PerComW.2014.6815206.

Ntakolia, C. *et al.* (2022) 'Machine learning applied on the district heating and cooling sector: a review', *Energy Systems*, 13(1), pp. 1–30. Available at: https://doi.org/10.1007/s12667-020-00405-9.

Rongali, S. *et al.* (2015) 'A context vector regression based approach for demand forecasting in district heating networks', in *2015 IEEE Innovative Smart Grid Technologies - Asia (ISGT ASIA)*. *2015 IEEE Innovative Smart Grid Technologies - Asia (ISGT ASIA)*, pp. 1–6. Available at: https://doi.org/10.1109/ISGT-Asia.2015.7387141.

Schaefer, A.M., Udluft, S. and Zimmermann, H.-G. (2008) 'Learning long-term dependencies with recurrent neural networks', *Neurocomputing*, 71(13), pp. 2481–2488. Available at: https://doi.org/10.1016/j.neucom.2007.12.036.

scikit-learn: machine learning in Python — scikit-learn 1.2.2 documentation (2023). Available at: https://scikit-learn.org/stable/ (Accessed: 12 June 2023).

sklearn.preprocessing.MinMaxScaler (no date) *scikit-learn*. Available at: https://scikit-learn/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html (Accessed: 8 June 2023).

Suryanarayana, G. *et al.* (2018) 'Thermal load forecasting in district heating networks using deep learning and advanced feature selection methods', *Energy*, 157, pp. 141–149. Available at: https://doi.org/10.1016/j.energy.2018.05.111.

TensorFlow (2023). Available at: https://www.tensorflow.org/ (Accessed: 12 June 2023).

Zhao, J., Li, J. and Shan, Y. (2021) 'Research on a forecasted load-and time delay-based model predictive control (MPC) district energy system model', *Energy and Buildings*, 231, p. 110631. Available at: https://doi.org/10.1016/j.enbuild.2020.110631.

Zimmerman, N., Kyprianidis, K. and Lindberg, C.-F. (2019) 'Achieving Lower District Heating Network Temperatures Using Feed-Forward MPC', *Materials*, 12(15), p. 2465. Available at: https://doi.org/10.3390/ma12152465.

Appendix

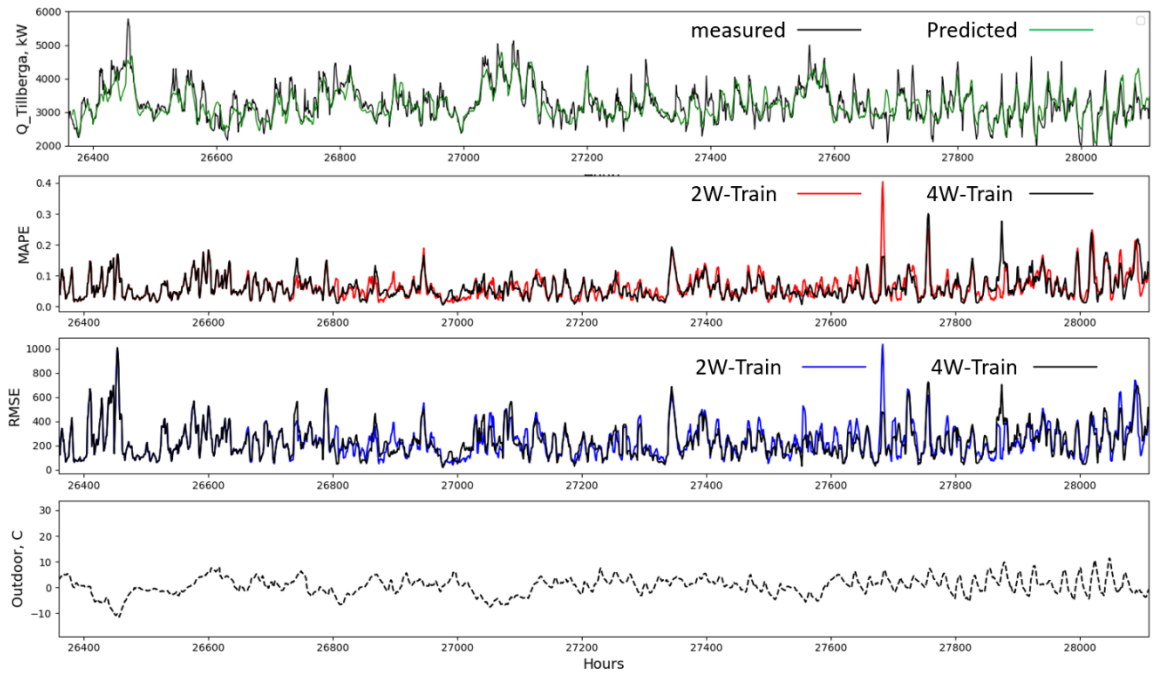


Figure 7: Forecasted heat load and Performance indicators (MAPE and RMSE) for **Tillberga** region with data from 1st Jan 2022 to 31st March 2022 with 6 **hr** prediction horizon

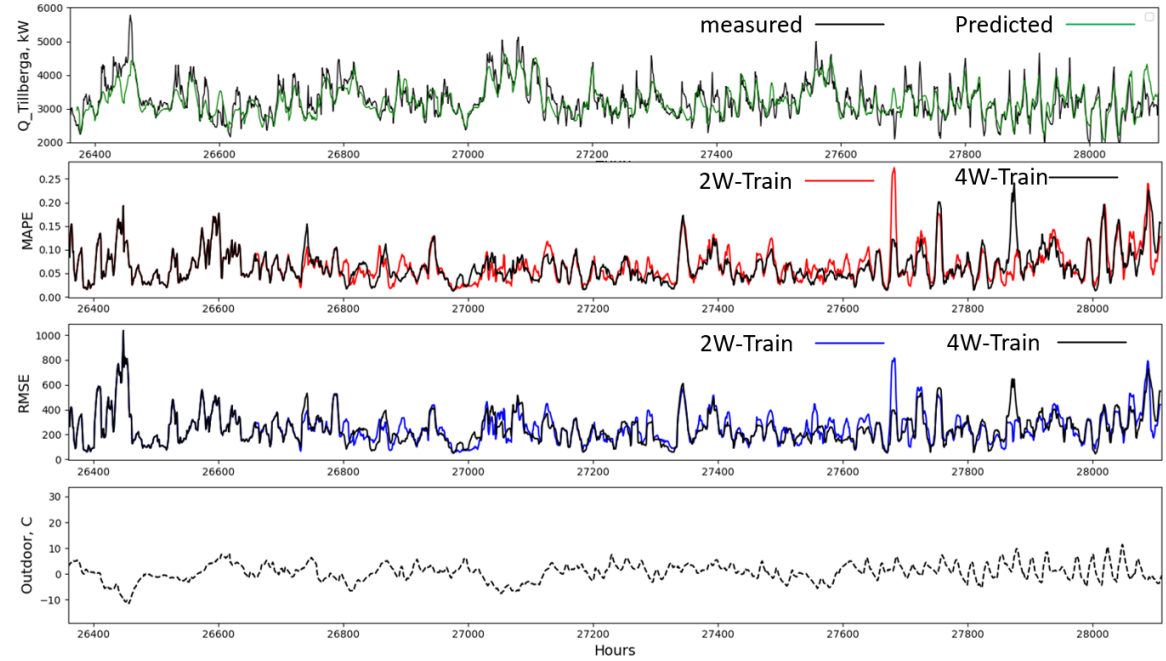


Figure 8: Forecasted heat load and Performance indicators (MAPE and RMSE) for **Tillberga** region with data from 1st Jan 2022 to 31st March 2022 with **10 hr** prediction horizon

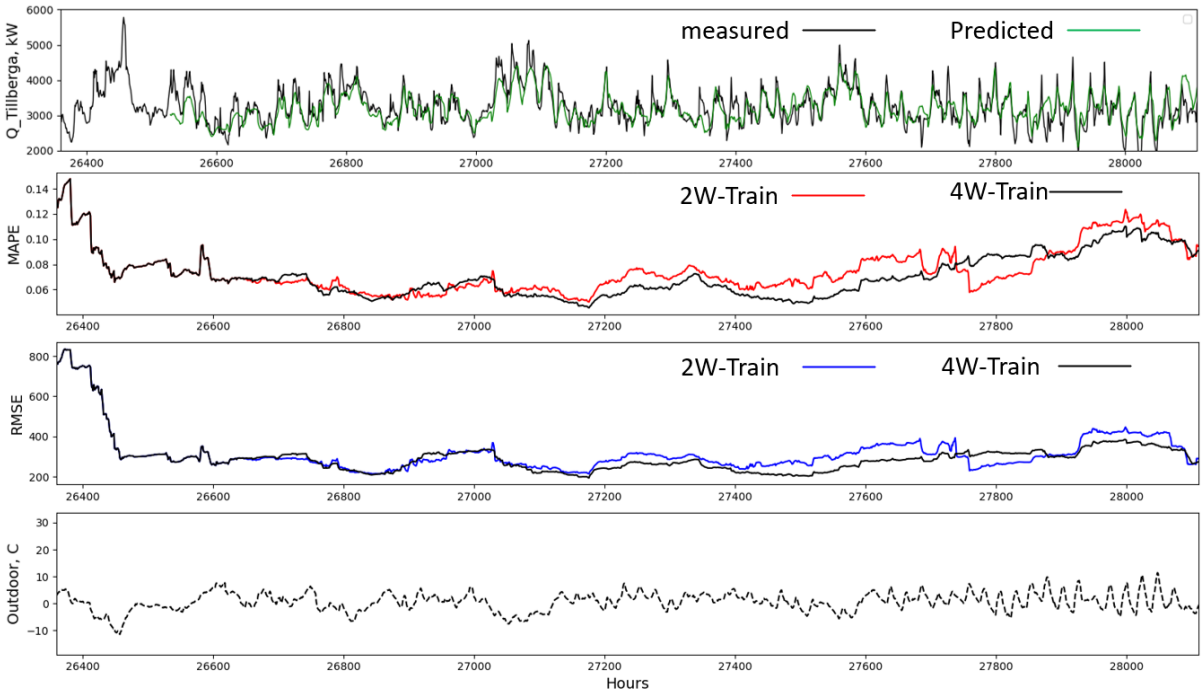


Figure 9: Forecasted heat load and Performance indicators (MAPE and RMSE) for **Tillberga** region with data from 1st Jan 2022 to 31st March 2022 with **1-week** prediction horizon

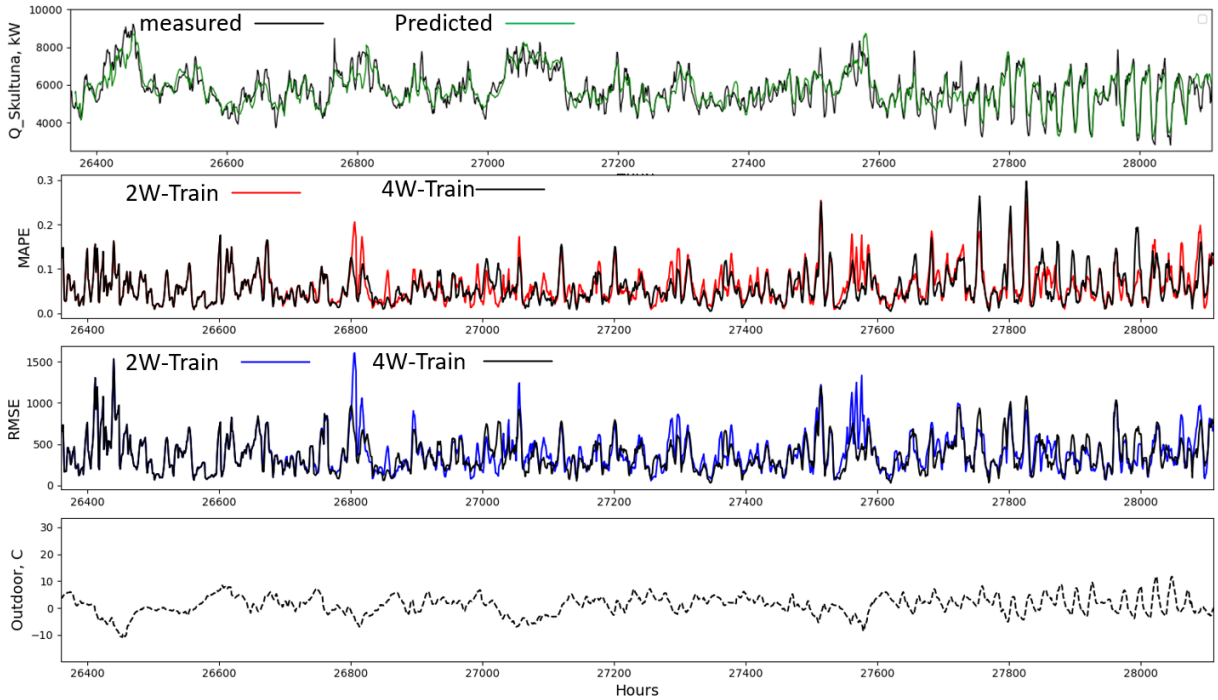


Figure 10: Forecasted heat load and Performance indicators (MAPE and RMSE) for **Skultuna** region with data from 1st Jan 2022 to 31st March 2022 with 6 **hr** prediction horizon

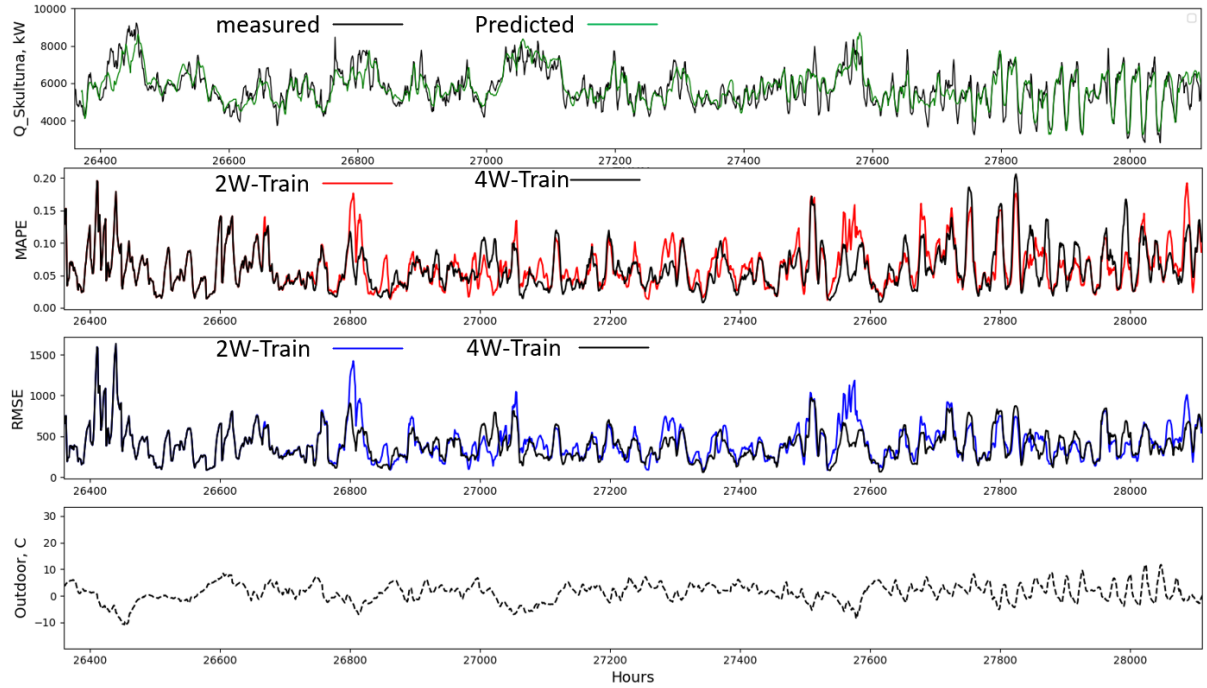


Figure 11: Forecasted heat load and Performance indicators (MAPE and RMSE) for **Skultuna** region with data from 1st Jan 2022 to 31st March 2022 with **10 hr** prediction horizon

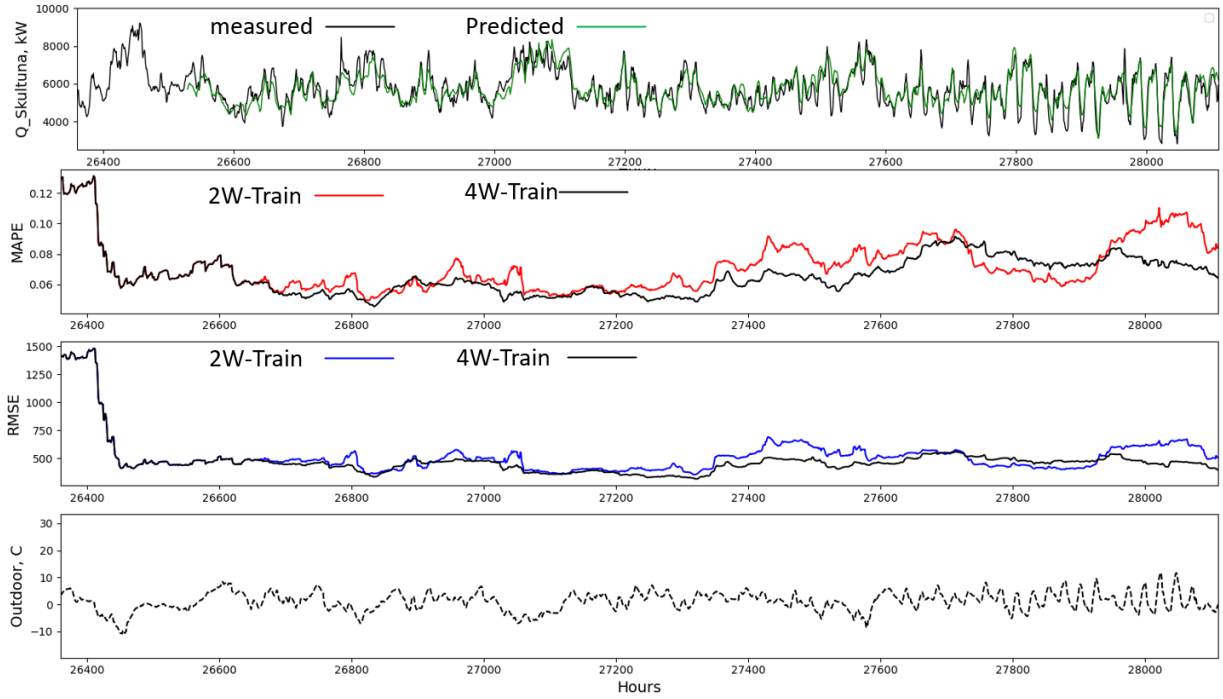


Figure 12: Forecasted heat load and Performance indicators (MAPE and RMSE) for **Skultuna** region with data from 1st Jan 2022 to 31st March 2022 with **1-week** prediction horizon

Retrofitting Biomass Combined Heat and Power Plant for Biofuel Production

Hao Chen [a,*], Daheem Mehmood [a], Erik Dahlquist [a], Konstantinos Kyprianidis [a]

[a] *School of Business, Society and Engineering, Mälardalen University, PO Box 883, SE-721 23 Västerås, Sweden*
hao.chen@mdu.se

Abstract

Thermochemical conversion processes of biomass, such as gasification and pyrolysis, can convert a wide range of feedstocks into liquid fuels, including forest residue, agricultural, food, and municipal solid waste. These more widely available and theoretically lower cost feedstocks make biofuel production through thermochemical pathway more cost-competitive. Furthermore, the thermochemical conversion pathway for biomass conversion could be relatively easy to integrate with the existing biomass combined heat and power plant, making it an attractive technology for the future commercialization of biofuel production through biomass. A detailed analysis was undertaken of a retrofitted biomass combined heat and power plant for biofuel production in this work. The biofuel production plant is designed to explore the polygeneration of hydrogen, biomethane, and bio-oil via the integration of gasification, pyrolysis, and renewable-powered electrolysis. The G-valve in the biomass circulating fluid bed plant, which is generally used for sand and char recycling, is retrofitted in the proposed system to fit the pyrolysis reaction for bio-oil production. Centering around the biomass circulating fluid bed gasifier, the system is also outfitted with a condensation and distillation process for bio-oil production, and a membrane reactor system for biomethane production. A mathematical model of the proposed biofuel production plant is established in Aspen Plus, followed by a performance investigation of the biofuel production plant under various design conditions. The limitations and opportunities of this retrofitted biomass combined heat and power plant for biofuel production are explored in this study.

1. Introduction

Biomass has been acknowledged as a premier renewable energy resource in the EU. Among various biomass conversion pathways, thermochemical conversion processes such as gasification and pyrolysis, are capable of producing biofuels from a diverse of feedstocks, including forest residuals, agricultural waste products, food waste, and municipal solid waste. These theoretically low-cost feedstocks make the thermochemical pathways of biomass conversion more economically viable for biofuels production [1].

As a thermochemical pathway for biomass conversion, fast pyrolysis has the potential to be integrated into existing biomass combined heat and power (CHP) plants [2], thereby enhancing its cost competitiveness for biofuel production. Karvonen et al. performed an environmental assessment on the integration of fast pyrolysis into a CHP plant [3]. This integration was achieved by using the heat from the char and non-condensing gas combustion to enhance heat and power generation in a CHP plant. The results indicated that the efficiency of stand-alone pyrolysis was improved from 59% to 71% upon integration into a CHP plant. A study on the integration of biomass fast pyrolysis with a municipal waste CHP plant was conducted by Kohl et al. [4]. The heat required in the biomass pyrolysis process was supplied by the hot flue gas from the CHP plant in this work, aiming at improving the pyrolysis product yield and retaining the district heat load simultaneously. It is noted that the operational hours of the CHP plant could be potentially increased by 57%, which makes this integration economically viable. Onarheim et al. performed a techno-economic analysis of a fast pyrolysis bio-oil production process with integration into an existing Fluid Bed Boiler CHP plant [5]. The sand heated in the CHP plant was sent to support the endothermic reaction in the pyrolysis reactor. Sensitivity analysis on different feedstocks and varying heat and electricity prices were also implemented in this study. The results showed that the economically advantages of this integration highly depend on the cost of heat and feedstocks. Zetterholm et al. completed a comprehensive evaluation of fast pyrolysis value chain configurations considering different types of locations, emissions, feedstocks, and final products [6]. The results showed that production cost for

crude pyrolysis liquid is in the range of 36-60 €/MWh (LHV of pyrolysis crude oil), and 61-90 €/MWh after further upgrading pyrolysis crude oil to diesel and petrol. It was also found that the integration of existing industrial infrastructure helps to mitigate the production cost.

Various gasification technologies are also considered in biofuel production processes since they provide excellent synergies. As a thermochemical process, biomass gasification can benefit from the integration of existing CHP plants. Piazzi et al. performed an experimental study to investigate the feasibility of retrofitting existing small-scale gasifier from CHP production to hydrogen and biofuel generation [7]. Co-production of syngas and biofuel by using the dual fluidized bed gasifier has been examined by Gustavsson et al., demonstrated a substantial enhancement on system efficiency [8]. An economic feasibility analysis of complementing existing CHP plants for hydrogen production was investigated by Naqvi et al., in this research the estimation cost for hydrogen production is 0.125-0.75 €/kg hydrogen [9]. Thunman et al. conducted an economic analysis of the GoBiGas plant, which is the first industrial installation for biomethane production with gasification [10]. This study found that the economic performance could be improved if integrated with existing infrastructure and low-grade feedstocks. Holmgren et al. examined the performance of gasification-based biofuel production systems with integration of district heating system [11]. It is concluded that the profitability of this system strongly depends on the specific production technologies and on the reference power production. The integration of existing CHP plant and gasification process for dimethyl ether or methanol production was analyzed by Salman et al. [12]. The results showed that the profitability could be notably improved by integrating gasification with CHP plants for biofuel production, as compared with the CHP plants that are only for heat and electricity generation.

In this study, a detailed analysis was undertaken of a retrofitted biomass combined heat and power plant for biofuel production. By retrofitting the G-valve in the biomass circulating fluid bed boiler for pyrolysis, the biofuel production process can benefit from the heat and hydrogen generation from the biomass gasification. It is expected that the proposed system could reduce the investment cost of biofuel production, and the integrated technologies could also serve as a solution for energy storage and transportation for renewables integration.

2. System description

The proposed pilot plant is designed to explore the polygeneration of hydrogen, biomethane, and bio-oil via the integration of biomass gasification, pyrolysis, and electrolysis with utilizing renewable energy. The primary component of the pilot plant is a Circulating Fluidized Boiler (CFB) with biomass as feedstock. The G-valve, typically used for sand and char recycling in the CFB, is retrofitted to fit the biomass pyrolysis reaction for biooil production. Centering around the Biomass CFB, the plant is also outfitted with cooling and distillation for bio-oil production, and membrane reactor system for biomethane production.

The schematic diagram of the facility is presented in Figure 1. During operation, ambient air is preheated to around 600 ℃ before being fed into the CFB, where the air transports and heats the feeding biomass up to enable the gasification process to occur downstream.

Then syngas generated from biomass gasification is separated from the solids (uncombusted biomass, char, and sand) in the cyclone. The solids, which still carry heat, are directed to the G-Valve (pyrolyzer), where the sensible heat of the solids is used to support the endothermic pyrolysis reaction and to generate pyrolysis vapor. After the cooling process within a condenser, the pyrolysis vapor becomes liquid bio-oil, which will be further upgraded to bio-gasoline or biodiesel in the hydrotreatment reactors with the presence of hydrogen. Meanwhile, the syngas after the cyclone will go through a reformer and a two-stage water-gas shift reactor to enhance the hydrogen generation. Finally, the existing carbon monoxide and carbon dioxide in the syngas, together with the generated hydrogen, is sent to a methanation reactor for biomethane generation.

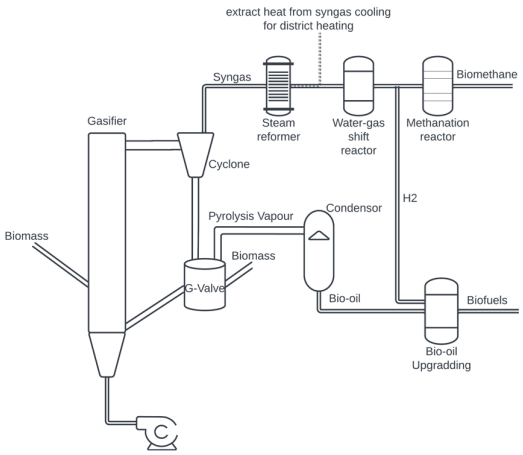


Figure 1, Schematic diagram of biomass pyrolysis, gasification, and electrolysis integrated polygeneration system

3. Process modeling

3.1 Biomass pyrolysis integrated with gasification

The process model for the entire biofuel production pilot plant was established in Aspen Plus to evaluate the system performance. Figure 2 illustrates the flowsheet of incorporating the pyrolysis process into the biomass CFB gasifier in Aspen Plus. The gasification was simulated by using two blocks, namely the DECOM block (RYield reactor) and the Gasifier block (Gibbs Reactor). Biomass is first converted into conventional components (C, H_2, O_2, N_2, S and ash) in the DECOM block, in which the product yield is calculated by an external Fortran code based on mass balance. The Gasifier block mixes the products from DECOM with air and simulates the gasification process by computing thermodynamic equilibrium.

A RYield reactor (Pyrolyzer block in Figure 2) is also used to conduct the pyrolysis process in the G-Valve. The mass yield fraction of pyrolysis product was taken from Lisa et al. [13] with the fixed pyrolysis temperature at 480 ℃. Char and ash generated from pyrolysis, along with recycled sand, are then directed to the Char Combustor block, where the solid char will be combusted. If needed, additional air will also be injected to the Biomass Comb block to supply heat for the pyrolysis. Part of the preheated air is also injected into the gasifier to support the endothermic gasification reaction. SiO_2 is used in this study to simulate sand in the Gasifier. The normalized feedstock ultimate analysis and the product yield for the pyrolysis reactor are given in Table 1. Peng Robinson cubic equation of state with the Boston-Mathias alpha function is used in Aspen Plus for all thermodynamic properties.

Table 1, Ultimate analysis of the feedstock and product yield for the pyrolysis

Ultimate analysis of the feedstock		Product yield for the pyrolysis	
Carbon	49.66 %	H_2	0.0000
Hydrogen	6.31 %	CO	0.0582
Oxygen	43.55 %	CO_2	0.0603
Nitrogen	0.10 %	CH_4	0.0028
Sulfur	0.08 %	C_2H_4	0.0028
Ash	0.30 %	Acetic Acid, $C_2H_4O_2$	0.1107
LHV	15.1 MJ/kg	Acetone, C_3H_6O	0.1272
		M-Cresol, C_7H_8O	0.0398
		Coniferyl Aldehyde, $C_{10}H_{10}O_3$	0.0068
		Guaiacol, $C_7H_8O_2$	0.2680
		Levoglucosan, $C_6H_{10}O_5$	0.0440
		Furfural, $C_5H_4O_2$	0.0294
		Water, H_2O	0.1480
		Char	0.0968

3.2 Bio-oil production and upgrading with onsite hydrogen generation

The pyrolysis vapor generated from the G-valve (Pyrolyzer) needs to be condensed to form bio-oil. To achieve this, a quench loop, depicted in Figure 3, is implemented to facilitate the condensation of the pyrolysis vapor into a liquid phase. The pyrolysis gas after the quench loop is sent back to the Char Combustion block (shown in Figure 2) to support the heat for gasification.

After the quench loop, bio-oil is separated from the aqueous phases in the pyrolysis product. To enhance the stability and heating value of the bio-oil, a hydrotreatment process is employed after the quench loop. The hydrotreatment reactions and operating parameters employed in the Hydrotreatment Reactor block are taken from Dutta et al. [14]. The product resulting from the hydrotreatment process is directed to the distillation column, where biofuel is separated out and produced. It is worth mentioning that the hydrogen required for the bio-oil upgrading is from gasification, which enables onsite self-sufficient hydrogen generation.

3.3 Bio-methane generation with renewables integration

As presented in Figure 4, to enhance the biofuel production of the pilot plant, syngas produced from the gasification process is mixed with the recycled gas from the bio-oil upgrading process and directed to the steam reformer to increase hydrogen production. To further increase hydrogen generation, a two-stage water-gas shift reactor (high temperature water-gas shift reactor, HT-WGS, and low temperature water-gas shift reactor, LT-WGS) is incorporated after the reformer. Subsequently, in the pressure swing adsorption (PSA) process, a portion of the hydrogen is diverted to the bio-oil upgrading process, while the remaining gas (primarily composed of H_2, CO, and CO_2) is compressed and channeled to the methanation reactor to produce bio-methane, aiming for enhanced biofuel production and carbon capture and utilization. Additionally, air preheating and high temperature steam generation are implemented into the process to improve the thermal efficiency of the entire pilot plant.

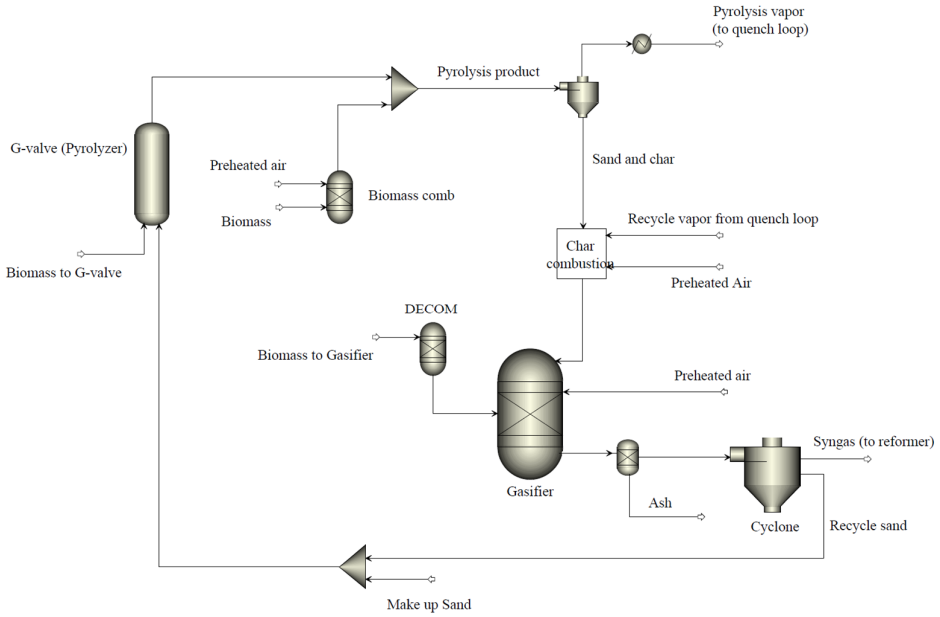


Figure 2, Process flowsheet of biomass pyrolysis integrated with gasification process in Aspen Plus

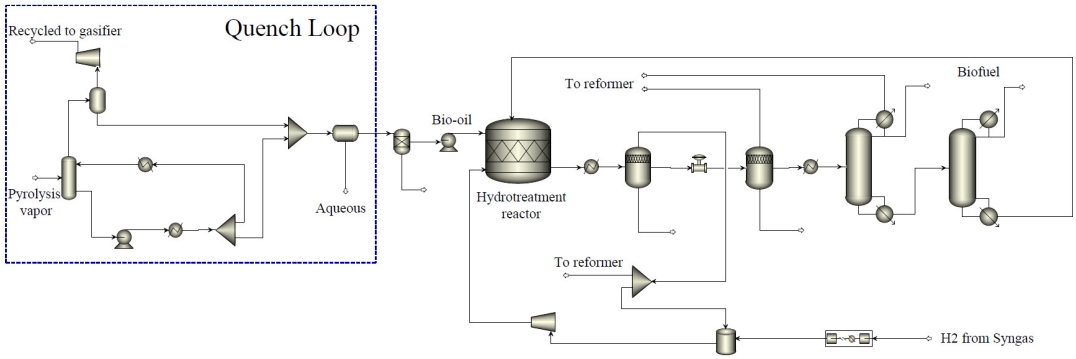


Figure 3, Process flowsheet of Bio-oil production and upgrading process in Aspen Plus

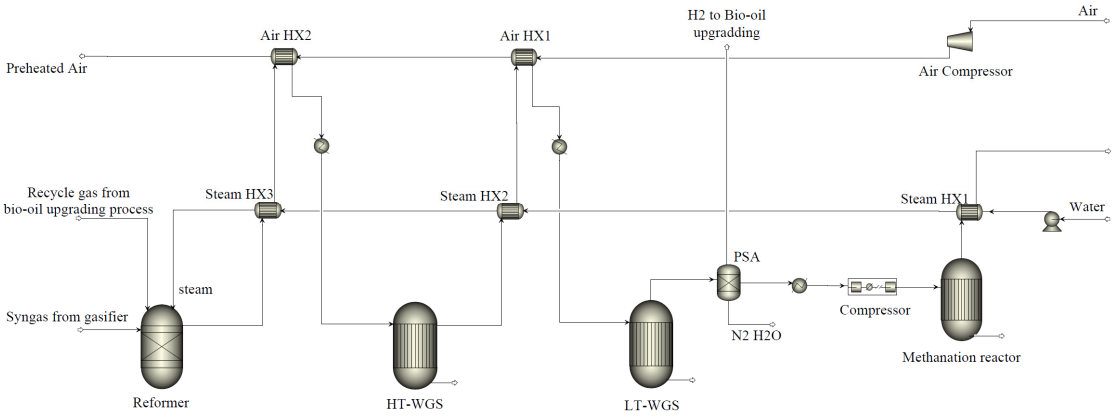


Figure 4, Process flowsheet of hydrogen and biomethane production process in Aspen Plus

4. Results

4.1 process modeling results

The goal of the process modeling is to determine the optimal parameters for the plant design to improve fuel production and profitability. In this baseline scenario, electrolyzers are not integrated in the polygeneration system. Based on the capacity of the pilot plant that is under construction at Malardalen University, the biomass feeding into the gasifier and pyrolyzer (G-Valve) are fixed at 45kg/hr and 15 kg/hr respectively,

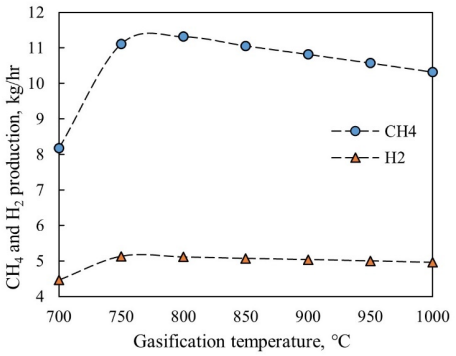


Figure 5, CH_4 production after the methanation reactor and H_2 production after LT-WGS reactor

As aforementioned, the pyrolysis and gasification process are coupled in the polygeneration plant by taking the heat from the recycling sand to support the endothermic pyrolysis process. The uncombusted solid left from the pyrolysis is then recycled back to the gasifier to participate in the gasification process. Therefore, the operating condition of the gasifier has a major impact on the downstream processes such as bio-oil production, hydrogen and biomethane generation. A sensitivity analysis is performed in this work to investigate the impacts of operating temperature of gasifier on hydrogen and methane production of the proposed system. Figure 5 shows the methane production (after methanation reactor) and hydrogen yield (after the LT-WGS reactor) when the gasification temperature varies from 700 to 1000 ℃. As illustrated in Figure 5, CH_4 and H_2 production increase when gasification temperature rises from 700 to 800 ℃, after which the CH_4 and H_2 yields start to drop if further increasing the gasification temperature from 800 to 1000 ℃. The reason behind the peak production of H_2 and CH_4 at 800 ℃ gasification temperature is that, while gasification benefits from the higher temperature, more air is required to support the higher gasification temperature through combustion, thus resulting in the reduced CO, H_2, and CH_4 in the syngas composition, and eventually causing the drop of H_2 and CH_4 production after WGS reactor and methanation reactor.

Figure 6 shows the gas composition change along with the process streams (after gasification, reforming, and two stage water-gas shift reaction), it can be seen that hydrogen composition in the syngas increased significantly after reforming and water-gas shift reaction.

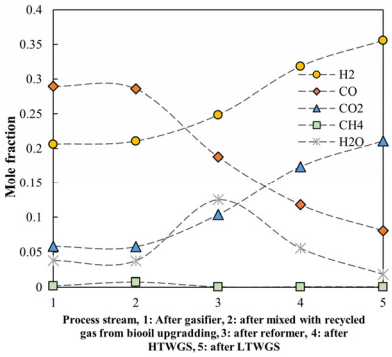


Figure 6, Mole concentration at varied location of the process line

4.2 Air, heat, and power consumption

Air and power consumption of the polygeneration system under varied operating gasification temperatures are summarized in Figure 7. As we discussed before, more air is injected into the gasifier to maintain a higher gasification temperature, which contributes to the increase in air consumption. As also shown in Figure 7a, in the case of gasification temperature lower than 800 ℃, the heat carried by the recycled sand is not enough to support the endothermic reaction, therefore air is feeding into the pyrolyzer to support the pyrolysis process. The same happens in the steam reformer reactor, air is also injected into the reformer to supply the heat (so called auto-thermal reforming). It is worth noting that the air required by the pyrolysis and reformer are calculated and controlled by an external Fortran code integrated in Aspen Plus.

Figure 7b demonstrates the power consumption of the polygeneration plant under different operating conditions. As shown in Figure 7b, more than half of the power consumption comes from the gas compression for the methanation. The methanation reactor operates at a high pressure above 30 bars, which consumes a large amount of power (about 20 kW, shown in Figure 7b) to pressurize the syngas before feeding into the methanation reactor. The second largest power consumption in the system is from the hydrogen compressor, as shown in Figure 7b. Hydrogen compressor is employed to compress the hydrogen (generated from the gasification, reforming, and water-gas shifting process) to the operating pressure (40 bar) of the bio-oil

hydrotreatment reactor, therefore resulting in an unneglectable proportion of power consumption.

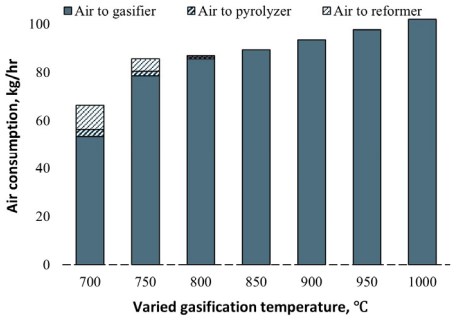


Figure 7a, Air consumption in the polygeneration system

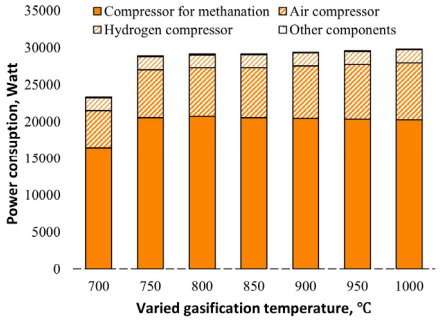


Figure 7b, Power consumption in the polygeneration system, baseline scenario

Heat requirements and the heat produced in the polygeneration plant are described in Figure 8. The heat demand comes mainly from the bio-oil upgrading process, such as hydrotreatment process, water separation and distillation process, about 2.8 kW. The operating conditions of the gasifier hardly affect the bio-oil production process, thus resulting in the nearly constant heat requirement with varied gasification temperatures, as shown in Figure 8a. Figure 8b shows the changes of the heat produced in the polygeneration plant, as more combustion is required to support the higher gasification temperature, more heat released from the system, around 9 - 10 kW, which provides an opportunity to be used for district heating.

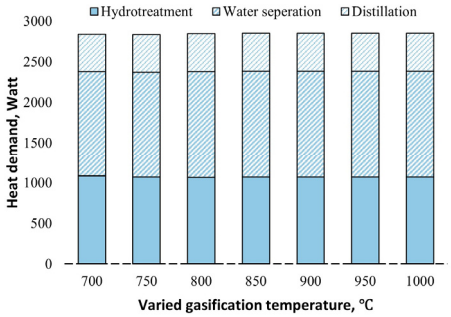


Figure 8a, Heat demand in the polygeneration system

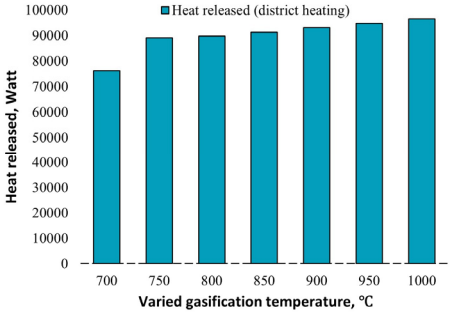


Figure 8b, Heat produced in the polygeneration system

4.3 Carbon efficiency

Figure 9 shows the carbon distribution and carbon efficiency of the polygeneration system under varied gasification temperatures. Carbon efficiency represents the proportion of carbon that has been converted into biofuels from the feedstock. As shown in Figure 9, around 40% of the carbon from the biomass could be captured in biomethane and bio-oil. The optimal gasification temperature in terms of the highest carbon efficiency (around 40%) is 800 ℃. And it is also worth noting that a large proportion of carbon is left in the ash when the operating temperature of gasifier is lower than 750 ℃, which is not favorable.

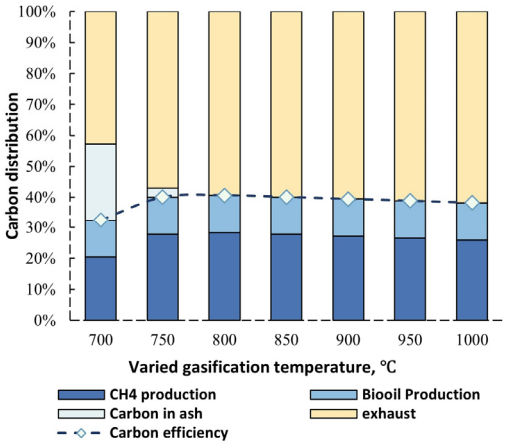


Figure 9, Carbon distribution in the final products and carbon efficiency of the polygeneration system

Figure 10 illustrates the carbon flows in the polygeneration system in the case of 800 ℃ gasification temperature. It is obvious that more than half of the carbon is still released into the atmosphere through waste and exhaust in this case.

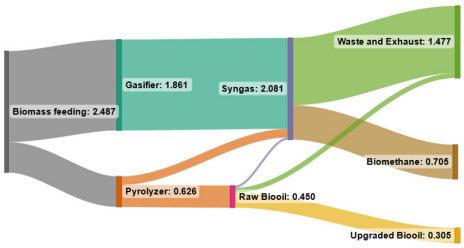


Figure 10, Carbon flows (kmol/hr) in the polygeneration system with 800 ℃ gasification temperature

Since the capacity of the pyrolyzer (G-Valve) could be relatively easy to scale up, a sensitivity analysis of pyrolyzer capacity (varied biomass feeding mass flowrate from 15 kg/hr to 45 kg/hr) is performed in this work. Table 2 summarizes the fuel productions and carbon efficiencies of the polygeneration system under varied pyrolyzer capacities, while the gasifier operating temperature and the biomass feeding flowrate in the gasifier are fixed at 800 ℃ and 45 kg/hr respectively in this case.

Table 2, Fuel productions and carbon efficiencies of the polygeneration system

Biomass to pyrolyzer (kg/hr)	Biooil production (kg/hr)	Lower heating value of biomethane (kWh/kg)	Biomethane production (kg/hr)	Carbon efficiency (%)
15	4.8	10.1	11.3	40.6
30	9.5	10.1	12.5	43.9
45	14.2	10.1	13.5	45.7

5. Summary and Discussions

A polygeneration system of retrofitting the existing biomass CHP plant for biofuel production was proposed and analyzed in this work. The process modeling of the polygeneration system, which integrates biomass gasification and pyrolysis to generate biofuels (biooil and biomethane), is performed in Aspen Plus. Sensitivity analysis of the key design parameters, such as gasification temperature, was conducted to investigate the impacts on system performance.

Retrofitting of existing CHP plant for biofuel production provides good opportunities for sustainable fuel generation and surplus renewable energy storage. By the integration of gasification and pyrolysis, the uncombusted char left from pyrolysis could be used to support the endothermic gasification process, and the hydrogen generated from gasification could be used to upgrade the biooil through hydrotreatment, thus improving the fuel production and profitability of such systems. The results also shows that the optimal gasification temperature in terms of enhancing biomethane and hydrogen production is 800 ℃. The carbon efficiency of the entire system could reach up to 40%.

It can be expected that when integrated with renewable energy, the polygeneration system could benefit from the oxygen and hydrogen produced by renewable-powered electrolysis, which could increase the biomethane production, this will be explored in our future work. It is also worth noting that a large amount of heat is produced in the polygeneration process, which could also be considered for district heating.

The results of this process modeling work will be utilized to optimize and guide the construction of a pilot scale reactor at Mälardalen University, Västerås, Sweden. Furthermore, a concurrent investigation into economic analysis is currently underway to explore the economic feasibility of such systems.

Acknowledgment

This work was carried out in IFAISTOS project (Intelligent electroFuel production for An Integrated STOrage System), which is funded under the framework of the joint programming initiative ERA-Net Smart Energy Systems. The initiative has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements no. 646039 and no. 755970.

References

[1] Van Dyk, S., Su, J., Mcmillan, J. D., & Saddler, J. (2019). Potential synergies of drop-in biofuel production with further co-processing at oil refineries. Biofuels, Bioproducts and Biorefining, 13(3), 760-775.

[2] Kohl, T., Laukkanen, T. P., & Järvinen, M. P. (2014). Integration of biomass fast pyrolysis and precedent feedstock steam drying with a municipal combined heat and power plant. Biomass and Bioenergy, 71, 413-430.

[3] Karvonen, J., Kunttu, J., Suominen, T., Kangas, J., Leskinen, P., & Judl, J. (2018). Integrating fast

pyrolysis reactor with combined heat and power plant improves environmental and energy efficiency in bio-oil production. Journal of Cleaner Production, 183, 143-152.
[4] Kohl, T., Teles, M., Melin, K., Laukkanen, T., Järvinen, M., Park, S. W., & Guidici, R. (2015). Exergoeconomic assessment of CHP-integrated biomass upgrading. Applied energy, 156, 290-305.
[5] Onarheim, K., Lehto, J., & Solantausta, Y. (2015). Technoeconomic assessment of a fast pyrolysis bio-oil production process integrated to a fluidized bed boiler. Energy & Fuels, 29(9), 5885-5893.
[6] Zetterholm, J., Wetterlund, E., Pettersson, K., & Lundgren, J. (2018). Evaluation of value chain configurations for fast pyrolysis of lignocellulosic biomass-Integration, feedstock, and product choice. Energy, 144, 564-575.
[7] S. Piazzi, L. Menin, D. Antolini, F. Patuzzi, and M. Baratieri, Piazzi, S., Menin, L., Antolini, D., Patuzzi, F., & Baratieri, M. (2021). Potential to retrofit existing small-scale gasifiers through steam gasification of biomass residues for hydrogen and biofuels production. International Journal of Hydrogen Energy, 46(13), 8972-8985.
[8] Gustavsson, C., & Hulteberg, C. (2016). Co-production of gasification based biofuels in existing combined heat and power plants–Analysis of production capacity and integration potential. Energy, 111, 830-840.
[9] Naqvi, M., Dahlquist, E., & Yan, J. (2017). Complementing existing CHP plants using biomass for production of hydrogen and burning the residual gas in a CHP boiler. Biofuels, 8(6), 675-683.
[10] Thunman, H., Gustavsson, C., Larsson, A., Gunnarsson, I., & Tengberg, F. (2019). Economic assessment of advanced biofuel production via gasification using cost data from the GoBiGas plant. Energy Science & Engineering, 7(1), 217-229.
[11] Holmgren, K. M., Berntsson, T. S., Andersson, E., & Rydberg, T. (2016). Comparison of integration options for gasification-based biofuel production systems–Economic and greenhouse gas emission implications. Energy, 111, 272-294.
[12] Salman, C. A., Naqvi, M., Thorin, E., & Yan, J. (2018). Gasification process integration with existing combined heat and power plants for polygeneration of dimethyl ether or methanol: A detailed profitability analysis. Applied Energy, 226, 116-128.
[13] Lisa, K., French, R. J., Orton, K. A., Dutta, A., & Schaidle, J. A. (2017). Production of low-oxygen bio-oil via ex situ catalytic fast pyrolysis and hydrotreating. Fuel, 207, 413-422.
[14] Dutta, A., Sahir, A., & Tan, E., (2015). Thermochemical Research Pathways with In Situ and Ex Situ Upgrading of Fast Pyrolysis Vapors, U.S. National Renewable Energy Laboratory, NREL/TP-5100 -62455.

Appendix

Table 3, Biooil composition

Biooil composition	Mass fraction, %	LHV, MJ/kg
Benzene	32.1	36.3
Ethanol	25.2	
Methylcyclohexane	11.6	
Cyclohexane	20.0	
Tetrahydrofurfuryl alcohol	9.7	
Ethylbenzene	1.4	

Table 4, Specifications used for reformer, water gas shift reactor, and methanation reactor

Block name		Specifications
Gasifier (RGibbs)	Pressure	2.02 bar
	Temperature	750 - 1000 ℃
	Calculation option	Calculate phase equilibrium and chemical equilibrium
Steam reformer (RGibbs)	Pressure	-0.20 bar
	Temperature	800 ℃
	Calculation option	Calculate phase equilibrium and chemical equilibrium
HT-WGS (REquil)	Pressure drop	-0.35 bar
	Inlet temperature	340 ℃
	Reactions	$CO + H_2O = CO_2 + H_2$
LT-WGS (REquil)	Pressure drop	-0.35 bar
	Inlet temperature	220 ℃
	Reactions	$CO + H_2O = CO_2 + H_2$
Methanation reactor (REquil)	Pressure	30 bar
	Temperature	360 ℃
	Reactions	$CO + H_2 = CH_4 + H_2O$
		$CO_2 + 4\ H_2 = CH_4 + 2H_2O$
		$CO + H_2O = CO_2 + H_2$

Table 5, Reactions and Operating parameters in the hydrotreatment reactor

		Operating parameters of the Hydrotreatment reactor
Temperature		400 ℃
Pressure		105 bar
Chemical reaction considered in the Hydrotreatment reactor		
Reaction number	Fractional conversion component and rate	Reactions
1	Acetic Acid, 1	Acetic Acid+ $2\ H_2$ = Ethanol+ H_2O
2	Furfural, 1	Furfural + $3\ H_2$ = Tetrahydrofurfuryl alcohol
3	Levoglucosan, 1	Levoglucosan + H_2 + H_2O = Sorbitol
4	M-Cresol, 0.26	H_2 + M-Cresol = Toluene + H_2O

5	M-Cresol, 1	4 H_2 + M-Cresol = Methylcyclohexane + H_2O
6	Guaiacol, 0.2	Guaiacol + 6 H_2 = Cyclohexane + 2 H_2O+ CH_4
7	Guaiacol, 1	Guaiacol + 3 H_2 = 2 H_2O+ CH_4+ Benzene
8	Benzene, 0.2	Benzene + 3 H_2 = Cyclohexane
9	Coniferyl Aldehyde, 0.5	Coniferyl Aldehyde + 2 H_2 = Toluene + 2 CO+ CH_4+ H_2O
10	Toluene, 1	Toluene + 3 H_2 = Methylcyclohexane
11	Coniferyl Aldehyde, 1	Coniferyl Aldehyde + 3 H_2 = Ethylbenzene + CO_2+ CH_4+ H_2O