Oil Production Forecasting with Uncertainty Description Using Data Driven Proxy Model

Javad Tavakolifaradonbe*, Ali Moradi, Britt M. E. Moldestad

Department of Process, Energy and Environmental Technology, University of South-Eastern Norway
h.javadtf@yahoo.com, ali.moradi@usn.no, britt.moldestad@usn.no

Abstract

The petroleum industry operates under great uncertainty. Achieving an efficient approach to quantify uncertainty in oil production models is of key importance in supporting decision-makers to find suitable strategies for mitigating risks and maximizing profit. Uncertainty quantification is commonly performed based on the Monte Carlo approach and this is a very time-consuming process by using the physics-based models developed by reservoir simulators. To solve this challenge, data-driven proxy models which are less complex and computationally efficient can be used as an alternative. This paper aims to investigate the functionality of the ANN method in developing proxy models for uncertainty quantification of oil production from advanced wells. The investigation is conducted through a case study for uncertainty assessment of cumulative oil and water productions from a long horizontal well with ICD completion and zonal isolation in a synthetic reservoir for 10 years. In this study, the Eclipse® reservoir simulator is used for developing the base case model and it is coupled with MATLAB® for generating the required data sets to train and test the ANN proxy model. According to the obtained results, the trained and developed ANN proxy model can predict the production of oil and water from advanced wells accurately with a mean error of less than 4%. Besides, the proxy model is 150 times faster than the Eclipse model and can solve the challenge of the time-consuming process of uncertainty quantification.

1. Introduction

Nowadays, to develop long-term oil production models, engineers need to predict the behavior of reservoirs by utilizing geological features [1]. However, the defining variables have a wide range and are many and various. Thereby integrating the parameters one by one cannot describe the reservoirs accurately. Hence some simplifications should be applied to solve the model's difficulties [1], [2]. First, the most impactful parameters should be determined. Sensitivity analysis is a good solution for figuring out the input parameters with the most likely influence [3]. By considering some assumptions on variables and their distribution, numerical models will be generated to simulate the reservoir. However, there is still an unsolved issue, namely the treatment of uncertainty.

Defeating the uncertainty in anticipating the production requires a huge number of simulations, which is unfeasible because of the long computation time. [4], [5]. Different techniques have been focused on, and among them, proxy models have received more attention. The first proxy model was innovated by utilizing a bilinear polynomial of inputs. These analytical models have been trained and developed to behave like simulators while consuming less time. In this way prediction, analysis, and finally optimization will be performed more efficiently [6], [7].

So far it can be mentioned that accuracy and acceleration are two important characteristics that should be considered during the reservoir simulations. Artificial Neural Networks (ANNs) have been introduced as a practical solution. In a study, ANNs were applied as a proxy model to assess the uncertainty in production prediction [8]. Another study investigated different architectures of the neural networks in reducing the time consumption of reservoir simulation [9]. Artificial neural network separately or in combination with the genetic algorithm was utilized to grab nonlinearities of problems [2], [10]. Apart from optimization of the algorithm, there are several studies related to the application of ANN in engineering and production. Shaik et al. [2] predicted the life time of a pipeline by applying ANN. Otchere et al. [7] forecasted the features of a petroleum reservoir by using supervised machine learning paradigms. Moreover, the application of neural networks in production prediction was also proposed by Yuan et al. [11]. Through all the previous studies it is mentioned that the quality and accuracy of a proxy model highly depend on the training step.

This study focuses on the applicability of ANN as a proxy model for assessing the uncertainty in production prediction. Based on the Design of Experiments (DOE), a set of reliable data is produced by coupling MATLAB and Eclipse. Then the proxy model is trained, and the trained model is used to assess the uncertainty based on the Monte Carlo sampling principle. The main purpose of this paper is to propagate a methodology to allow for a more reliable decision about the productivity of a reservoir based on geological parameters.

1.1. Definition of Artificial Neural Network

A neural network is simply a set of neurons that are connected. Each neuron, as it is depicted in Fig. 1, takes one or more inputs and gives an output based on defined functions. These functions may be Relu, Sigmoid, Gaussian, or the sign functions. It should be mentioned that synaptic weight is considered for different neurons [4].

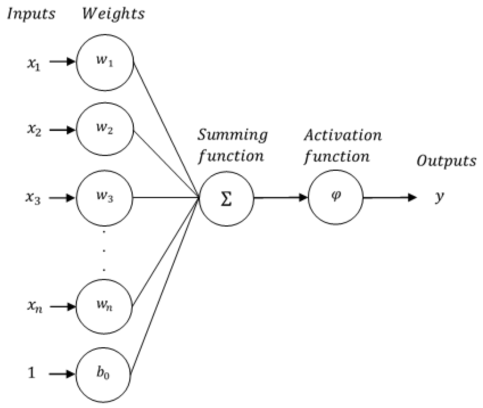

Figure 1: Schematic of a artificial neuron.

According to Fig. 1, x represents inputs, w is allocated to weight vector and b_0 is bias. Mathematically each neuron should be displayed by its function (f). So, if a neuron receives n input it would be represented as bellow:

$$f: R^{n+1} \times R^n \rightarrow R \qquad (1)$$

Satisfying

$$1: g: R \rightarrow R$$

$$2: W \in R^{n+1}, W = (w_1, w_2, \dots, w_3, b_0)$$

$$3: \forall\, x \in R^n, f(W, x) = g(\sum_{i=1}^{n} w_i z_i + b_0)$$

$$\mathrm{x} = (z_1, \dots, z_n) \qquad (2)$$

Where g is the transfer function. This basic function could model the higher-order functions by utilizing the collective behavior of a set of neurons which is called a layer. Indeed, a network is made of multilayer, consisting of an input layer, one or more intermediate or hidden layers, and an output layer, while each layer is composed of several neurons (Fig. 2).

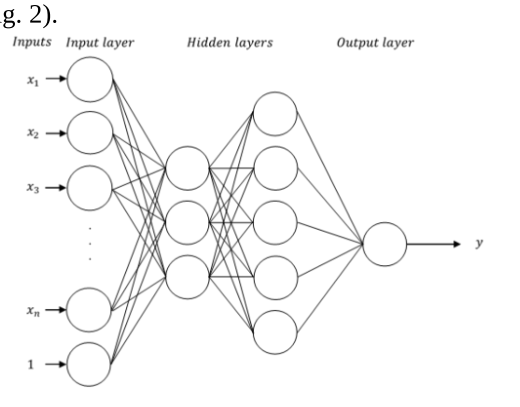

Figure 2: Schematic diagram of a neural network with two hidden layers.

Learning is the most important ability of a neural network, and a neural network will be able to generalize, classify and foresee [4], [12]. In other words, because of having experienced, neural networks will have recognition ability. But according to the learning, networks are divided into two classes, supervised and unsupervised networks. In a supervised learning class, input and output are fed into the network at the same time [7]. Then, the machine will learn how to reconfigure itself. On the other side, under unsupervised learning, the proxy is exposed to unlabeled input solely for clustering or comparing. It should be mentioned that to make the quadradic error of output at least, backpropagation is utilized in the supervised learning network, indeed it is a method of more accurate weight calculation [13].

Generally, based on the structure of the network and the operation of neurons, neural networks carry out a quite simple differentiable function. Indeed, after the learning phase and stabilizing the weight, the machine as a black box forecasts the phenomenon for new inputs [9]. Despite all of these, there are still some deterrents against utilizing ANNs. In other words, configuring the architecture of ANNs, namely the number of layers and number of neurons in each layer, should be found, while there is not any identification of a better architecture [5].

2. Methodology

2.1. Creating Proxy Model

Fig. 3 represents the steps of an algorithm to model a proxy. The Data sets and the previously proxy algorithms are considered as the most important factors for qualifying a proxy model. To make sure that all aspects of the model are dealt with, an infinite size dataset is required, which is practically impossible [3], [9], [12]. Some techniques of experimental design are enlisted to extract the utmost information with the least simulations.

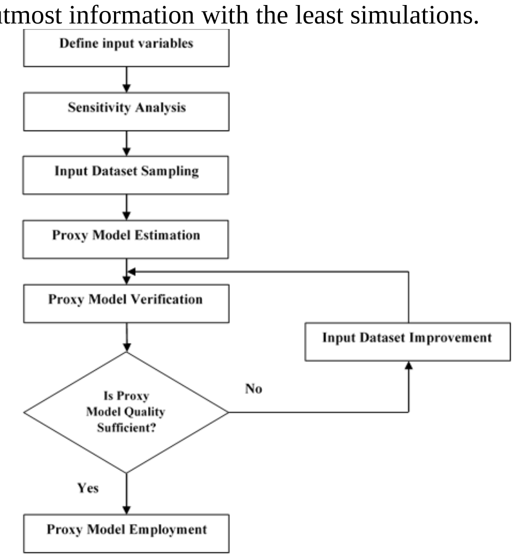

Figure 3: Schematic proxy model development [14].

Selecting the input variable highly depends on the type of problem and the level of knowledge of the project. It is recommended to consider all input variables at the beginning and then omit unimportant parameters through the sensitivity analysis step. Indeed, sensitivity analysis filters out less significant parameters on a simulation model. Consequently, an appropriate dataset will be prepared.

The accuracy of the proxy model highly originates from training. By utilizing a decent dataset sampling, the proxy model will be trained appropriately, and in the following, the estimation will be performed accurately. Thereby, verifying the model which is based on prediction accuracy will be satisfied [14].

2.1.1. Architecture of Artificial Neural Network

A neural network is considered a good proxy when it predicted a new case with acceptable error. Therefore, evaluation of the model should be performed sequentially to avoid overtraining. In this way, cross-validation, as one of the most popular methods, lets defined architecture stop learning when a validation error is raised [11], [12].

Overtraining also comes from a poorly structured network. Thereby identifying the appropriate number of hidden layers and their neurons is so requisite [1]. Moreover, the complexity of neural networks should also be limited. For this purpose, after finishing the learning phase, the pruning method will eliminate the connections with the smallest effect on the output error.

To obtain an optimal neural network for the defined method, both pruning and cross-validation were utilized.

2.2. Development of The Physic-based Model With Uncertainity Description

2.2.1. Defining Uncertain Input Domains

This study is conducted through modeling and forecasting oil and water production from an advanced horizontal well in a synthetic reservoir with uncertain properties for 10 years. The reservoir properties are the model inputs, and it is assumed that the value of some of the properties is uncertain. The uncertain reservoir parameters with their uncertainty range are reported in Tab. 1.

Table 1: Uncertain reservoir properties with their range.

Parameter	**Min**	**Mean**	**Max**
Porosity	0.15	0.23	0.27
Permeability in x-dir. [mD]	200	500	1000
Permeability in y-dir. [mD]	150	600	1200
Permeability in z-dir. [mD]	20	100	500
Irreducible water sat.	0.1	0.15	0.2
Residual oil saturation	0.05	0.1	0.15
Max. rel. perm. of water	0.2	0.4	0.5
Max. rel. perm. of oil	0.85	0.95	1
Initial water saturation	0.12	0.2	0.25
Capillary pressure [bar]	4	2.7	2
Aquifer prod. Index [m^3/d/bar]	2000	10000	15000

2.2.2. Determining the Most Impactful Uncertain Parameters

Uncertainty quantification based on the Monte Carlo approach requires many simulations. For each simulation run, a random combination of model input values is chosen, and the corresponding model outputs are calculated by using the simulator. This is a very time-consuming process when the system has several inputs. By filtering the less important inputs out and focusing on the most impactful input variables on the accuracy of the models, a bit of prediction accuracy is sacrificed but the speed of uncertainty assessment highly increases. The sensitivity analysis assesses the contribution of the uncertainty of each model input to the accuracy of the model outcomes and identifies the most important parameters of the system. Cumulative oil and water production are the most important outputs of oil models and are the model outputs in this paper. By performing sensitivity analysis on the uncertain reservoir parameters given in Tab. 1, the sensitivity coefficient of each reservoir parameter for the cumulative oil and water production is calculated. The obtained results are depicted as a tornado diagram in Fig. 4. Based on the presented results, the five most important input variables for predicting oil and water production are determined and given in Tab. 2. These input variables are the model inputs for the proxy model development and uncertainty quantification.

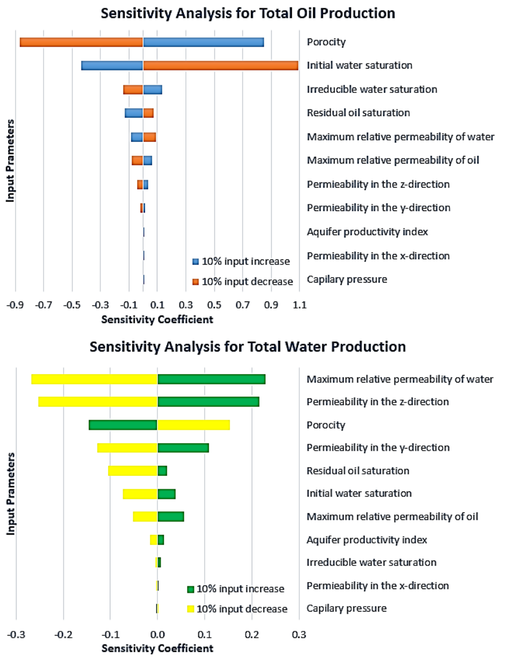

Figure 4: Sensitivity analysis of uncertain reservoir parameters.

Table 2: Uncertain input variables for uncertainty assessment.

Parameter	Min.	Mean	Max.
Porosity	0.15	0.23	0.27
Irreducible water saturation	0.1	0.15	0.2
Initial water saturation	0.12	0.2	0.25
Permeability in z-dir	20	100	500
Max. rel. perm. of water	0.2	0.4	0.5

2.2.3. Development of The Physics-based Model in Eclipse.

In this paper, the production prediction and uncertainty assessment are performed for primary oil production from a medium oil reservoir with a water drive. The reservoir fluid properties, as well as the temperature and pressure of the reservoir, are given in Tab. 3.

Table 3: Reservoir characteristics and fluid properties.

Parameter	Value
Oil density and Viscosity	900 kg/m^3, 2.5 cP
Water density and Viscosity	1050 kg/m^3, 0.45 cP
Gas-oil ratio (GOR)	50
Temperature and Pressure	60 ˚C, 200 bar

It is assumed that oil is produced from the reservoir near an advanced horizontal well with a length of 1000 m completed with Inflow Control Devices (ICDs) and zonal isolation. The diameters of the wellbore, the production tubing, and the ICDs are 8.5 inches, 5.5 inches, and 0.01 m respectively. The thickness and width of the reservoir are assumed to be 30 m and 70 m respectively. It is also assumed that the well is located 5.5 m below the top of the drainage area. The schematic of the near-well reservoir is shown in Fig. 5.

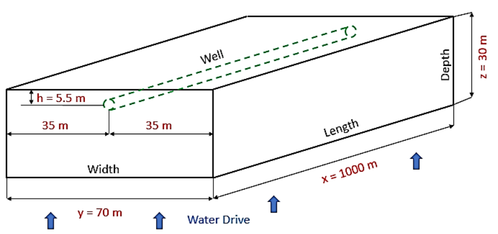

Figure 5: Schematic of the near-well reservoir.

To achieve a suitable grid setup, in the Y and Z directions finer meshes have been set near the wellbore and uniform meshes are considered in the X-direction. It is assumed that the horizontal well has 8 equivalent joints, each 125 m long. As a result, 8 uniform cells are considered for the reservoir in the X-direction. The grid resolution in Y and Z directions is illustrated in Fig. 6.

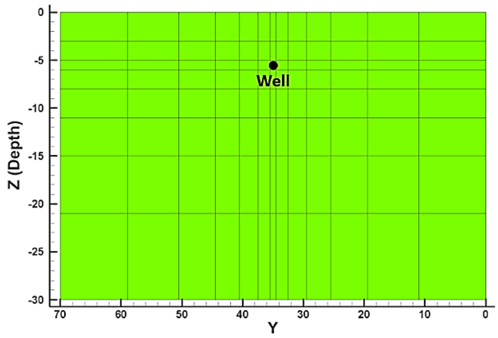

Figure 6: Grid resolution in the Y-Z plane.

In this study, Eclipse® which is a robust physics-based simulator is applied as a simulation tool. Due to the high pressure and low temperature of the reservoir, the reservoir condition is located well to the left-hand side of the critical point, and the black-oil model can be used for modeling fluid flow from the reservoir to the production tubing. Moreover, the multisegmented well model in the Eclipse simulator is used for developing the well model with ICD completion and zonal isolation. The well is considered to be controlled by the Bottom Hole Pressure (BHP), and the BHP is assumed to be 190 bar. Based on the mean value of the reservoir properties and the mentioned considerations and assumptions, a base model is developed in the Eclipse simulator. The cumulative oil and water production based on the base model for 10 years is shown in Fig. 7.

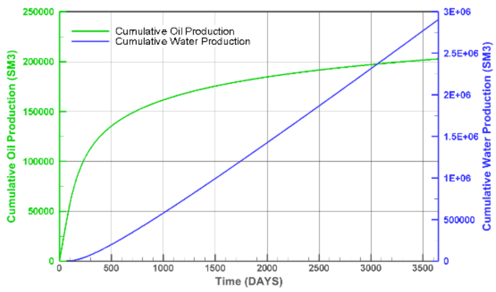

Figure 7: Base model water and oil production prediction for 10 years.

3. Results

3.1. Training and Test the ANN

The training data sets were generated by the Eclipse reservoir simulator. Inputs consisted of 5 variables and each variable accounted for 5 values, which gives the data sets with a size of 5^5. For each variable, the min. value, max. value, mean value, a value between min. and mean, and a value between mean and max. were opted. In addition, the main dataset accounted for 67% training, 25% validation, and 8% test.

Before feeding inputs to the machine for training, all values were normalized between [0,1] based on min value and max value. Outputs in the dataset accounted for accumulative oil and water production through 10 years. Because of that, for predicting fluid production of each year, fluid production of the previous year was considered as input too, Fig. 8.

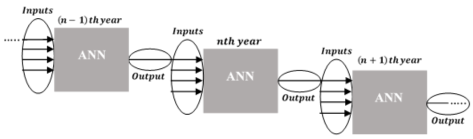

Figure 8: Schematic flow chart of predicting in sequential years.

For testing the proxy, the values of inputs opted ununiformly in a way that the machine had never experienced, although these values were between the minimum and the maximum values.

During testing the machine, prediction errors were calculated point by point within 10 years and for each state. State refers to the specific set of an input.

$$Error = \frac{simulation - prediction}{simulation} \qquad (3)$$

The machine predicted the oil production better than the water production. The mean error values are presented as a horizontal line in Fig. 9, 0.96 and 3.79 for oil and water respectively.

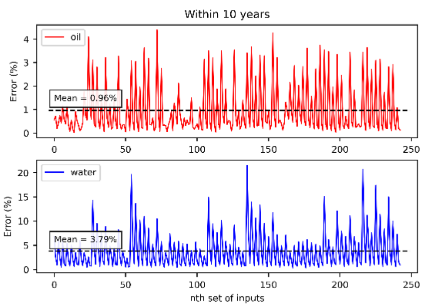

Figure 9: Prediction error for testing the proxy model.

3.2. Uncertainty Assessment in Production Prediction

Input datasets play a great role in studying uncertainty in production prediction. In addition, the uncertainty of input parameters makes it necessary to use a probabilistic approach [4], [6]. Although there are different methods for probabilistic data sampling, Latin Hypercubic Sampling (LHS) was chosen as an efficient sampling method [14]. Indeed, by considering the min and max of 5 effective parameters, the 100000 most probable sets of inputs were extracted according to the LHS approach. In this way, P (10), P (50), and P (90) are to be predicted.

After performing 100000 predictions for each time step. Cumulative probability distribution and probability density for accumulative oil and water production whithin 10th year is shown in Fig. 10, 11, 12 and 13.

Fig. 10 depicts a normal distribution for accumulative oil production. On the other side, Fig. 12 shows a lag normal distribution of water production.

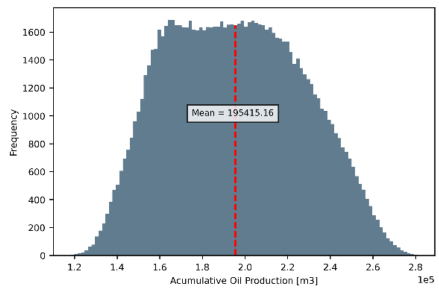

Figure 10: Probability distributionof cumulative oil production after 10 years.

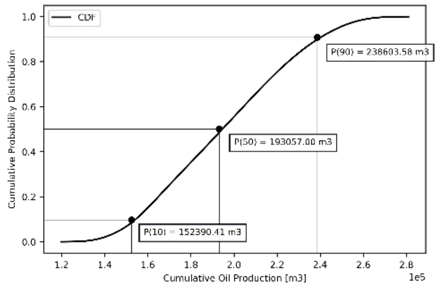

Figure 11: Cumulative probability distributiondiagram for predicting oil production after 10 years.

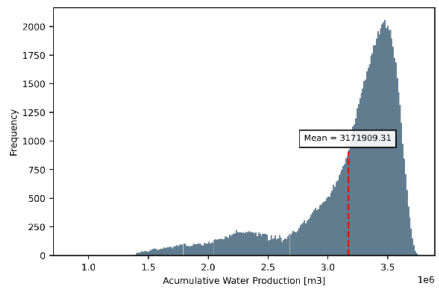

Figure 12: Probability distribution of cumulative water production after 10 years.

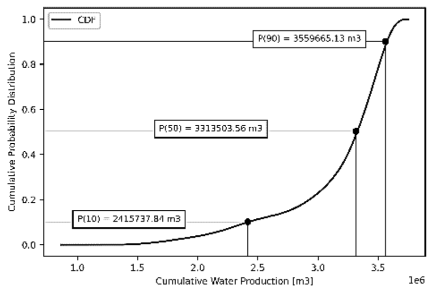

Figure 13: Cumulative probability distributiondiagram for predicting water production after 10 years.

As it is presented in Tab. 4, there is a 90% chance to produce less than 2.4E+05 m^3 oil and less than 3.6E+06 m^3 water after 10 years. In addition, cumulative oil and water production will be less than 1.5E+05 m^3 and 2.4E+06 m^3 respectively with a chance of 10%. In this case, the best estimation (P50) for cumulative oil and water production is 1.9E+05 m^3 and 3.3E+06 m^3 respectively.

Table 4: Summary of uncertainty results.

	P (10)	**P (50)**	**P (90)**
Oil	1.5E+05	1.9E+05	2.4E+05
Water	2.4E+06	3.3E+06	3.6E+06

4. Conclusion

As the oil reservoirs are developing, uncertainty assessment has become a priority. An unreliable prediction of the producibility of a reservoir may cause investment bankruptcy. Although this type of reservoir engineering problem could be solved by applying a reservoir simulator, the time-consuming process is a significant deterrent. Therefore, an artificial neural network with a stochastic approach has been enlisted to analyze uncertainty.

This paper presents an appropriate methodology to deal with assessing the uncertainty during the fluid production prediction. The paper compiles MATLAB and Eclipse and built an amendable optimal Neural Network by utilizing mathematical procedures, to reduce the time consumption of data extraction and prediction of an oil reservoir.

The results show that an appropriate proxy model can predict the production with an acceptable error of less than 4%. In addition, when utilizing ANN, the time consumption was reduced by a ratio of 150 times. It is also concluded that by increasing the size of a dataset, the time-consumption effectiveness of ANN will raise.

References

[1] J. Navrátil, A. King, J. Rios, G. Kollias, R. Torrado, and A. Codas, "Accelerating Physics-Based Simulations Using End-to-End Neural Network Proxies: An Application in Oil Reservoir Modeling," *Frontiers in Big Data*, vol. 2, p. 33, 2019, doi: 10.3389/fdata.2019.00033.

[2] T. Foroud, A. Seifi, and B. AminShahidi, "Assisted history matching using artificial neural network based global optimization method - Applications to Brugge field and a fractured Iranian reservoir," *Journal of Petroleum Science and Engineering*, vol. 123, 2014, doi: 10.1016/j.petrol.2014.07.034.

[3] A. F. Teixeira and A. R. Secchi, "Machine learning models to support reservoir production optimization," *IFAC-PapersOnLine*, vol. 52, no. 1, pp. 498–501, 2019, doi: 10.1016/j.ifacol.2019.06.111.

[4] D. Guérillot and J. Bruyelle, "Uncertainty assessment of hydrocarbon in place with geological models satisfying gravity-capillary equilibrium," *Society of Petroleum Engineers - SPE Kuwait Oil and Gas Show and Conference 2019, KOGS 2019*, 2019, doi: 10.2118/198099-ms.

[5] D. D. Monteiro *et al.*, "Uncertainty analysis for production forecast in oil wells," *SPE Latin American and Caribbean Petroleum Engineering Conference Proceedings*, 2017, doi: 10.2118/185550-ms.

[6] C. Wei *et al.*, "Uncertainty assessment in production forecasting and optimization for a giant multi-layered sandstone reservoir using optimized artificial neural network technology," *Society of Petroleum Engineers - SPE Reservoir Characterisation and Simulation Conference and Exhibition, RCSC 2017*, pp. 1179–1188, 2017, doi: 10.2118/186069-ms.

[7] D. A. Otchere, T. O. Arbi Ganat, R. Gholami, and S. Ridha, "Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties: Comparative analysis of ANN and SVM models," *Journal of Petroleum Science and Engineering*, vol. 200. 2021. doi: 10.1016/j.petrol.2020.108182.

[8] D. R. Guerillot and J. Bruyelle, "History Matching Methodology Using an Optimal Neural Network Proxy and a Global Optimization Method," vol. 2016, no. 1, pp. 1–5, 2016, doi: https://doi.org/10.3997/2214-4609.201602403.

[9] P. C. Silva, C. Maschio, and D. J. Schiozer, "Use of Neuro-Simulation techniques as proxies to reservoir simulator: Application in production history matching," *Journal of Petroleum Science and Engineering*, vol. 57, no. 3, pp. 273–280, 2007, doi: https://doi.org/10.1016/j.petrol.2006.10.012.

[10] D. R. Guérillot and J. Bruyelle, "Uncertainty assessment in production forecast with an optimal artificial neural network," in *SPE Middle East Oil and Gas Show and Conference, MEOS, Proceedings*, 2017, vol. 2017-March. doi: 10.2118/183921-ms.

[11] Z. Yuan, H. Huang, Y. Jiang, and J. Li, "Hybrid deep neural networks for reservoir production prediction," *Journal of Petroleum Science and Engineering*, vol. 197, 2021, doi: 10.1016/j.petrol.2020.108111.

[12] N. B. Shaik, S. R. Pedapati, and F. A. B A Dzubir, "Remaining useful life prediction of a piping system using artificial neural networks: A case study," *Ain Shams Engineering Journal*, 2021, doi: https://doi.org/10.1016/j.asej.2021.06.021.

[13] L. A. N. Costa, C. Maschio, and D. Schiozer, "Study of The Influece of Training Data set in Artificail Neural Network Applied to The History Matching Process," 2010.

[14] D. I. Zubarev, "Pros and cons of applying proxy-models as a substitute for full reservoir simulations," *Proceedings - SPE Annual Technical Conference and Exhibition*, vol. 5, pp. 3234–3256, 2009, doi: 10.2118/124815-ms.