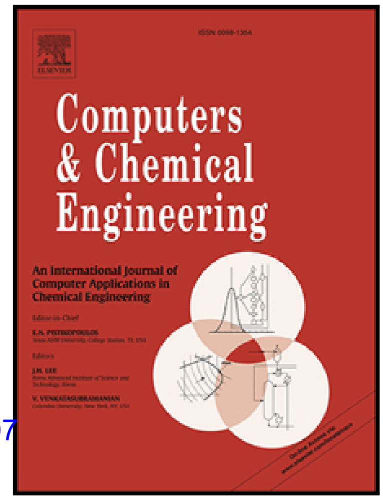
Journal Pre-proof



A Survey on the Application of Machine Learning and Metaheuristic Algorithms for Intelligent Proxy Modeling in Reservoir Simulation

Cuthbert Shang Wui Ng , Menad Nait Amar ,
Ashkan Jahanbani Ghahfarokhi , Lars Struen Imsland

PII: S0098-1354(22)00440-9
DOI: https://doi.org/10.1016/j.compchemeng.2022.108107
Reference: CACE 108107

To appear in: *Computers and Chemical Engineering*

Received date: 25 May 2022
Revised date: 12 December 2022
Accepted date: 15 December 2022

Please cite this article as: Cuthbert Shang Wui Ng , Menad Nait Amar , Ashkan Jahanbani Ghahfarokhi , Lars Struen Imsland , A Survey on the Application of Machine Learning and Metaheuristic Algorithms for Intelligent Proxy Modeling in Reservoir Simulation, *Computers and Chemical Engineering* (2022), doi: https://doi.org/10.1016/j.compchemeng.2022.108107

Highlights

- A survey of applying ML and coupled ML-metaheuristic paradigm in intelligent proxy modeling of NRS is provided.
- Discussion about relevant previous works is outlined.
- General methodology of developing models using ML is expounded.
- Application of ML and coupled ML-metaheuristic paradigm in different domains of reservoir simulation is discussed.

Paper Draft

A Survey on the Application of Machine Learning and Metaheuristic Algorithms for Intelligent Proxy Modeling in Reservoir Simulation

Cuthbert Shang Wui Ng[a,*], Menad Nait Amar[b], Ashkan Jahanbani Ghahfarokhi[a], Lars Struen Imsland[c]

[a]Department of Geoscience and Petroleum, Norwegian University of Science and Technology, Trondheim, Norway

[b]Département Etudes Thermodynamiques, Division Laboratoires, Sonatrach, Boumerdes, Algeria

[c]Department of Engineering Cybernetics, Norwegian University of Science and Technology, Trondheim, Norway

[*]Corresponding author.
Department of Geoscience and Petroleum, Norwegian University of Science and Technology, S. P. Andersens veg 15a, 7031 Trondheim, Norway.
cuthbert.s.w.ng@ntnu.no,

Abstract

Machine Learning (ML) has demonstrated its immense contribution to reservoir engineering, particularly reservoir simulation. The coupling of ML and metaheuristic algorithms illustrates huge potential for application in reservoir simulation, specifically in developing proxy models for fast reservoir simulation and optimization studies. This is conveniently termed the coupled ML-metaheuristic paradigm. Generally, proxy modeling has been extensively researched due to the expensive computational effort needed by traditional Numerical Reservoir Simulation (NRS). ML and the abovementioned coupled paradigm are effective in establishing proxy models. We conduct a survey on the employment of ML and the coupled paradigm in proxy modeling of NRS. We present the respective successful applications as reported in the literature. The benefits and limitations of these methods in intelligent proxy modeling are briefly explained. We opine that some study areas, including sampling techniques and dimensionality reduction methods, are worth investigating as part of the future research development of this technology.

Keywords

Machine Learning; Metaheuristic Algorithms; Data-Driven Modeling; Intelligent Proxies; Reservoir Engineering; Numerical Reservoir Simulation

1. INTRODUCTION

As global technology advances, the energy demand continues to evolve exponentially (Tillerson, 2008). This noticeable demand has made fossil fuels the dominant link in the energy subject area despite the continuous efforts made by the industrial sector to promote the vision and importance of renewable energies (British Petroleum, 2021). This source of energy, i.e., fossil fuels mainly from oil and gas reservoirs, goes through a step-by-step process to achieve the most desirable recovery factors. As a result, exploitation and development methods have been distinguished and classified into three categories, namely primary, secondary and tertiary recovery techniques (Ahmed, 2018). Fundamentally, these two latter techniques are designed to ensure the continuous production of hydrocarbons given the inefficacy of primary recovery. During primary recovery, the driving mechanism of hydrocarbon production originates from the natural source of energy associated with the rock and fluids in the reservoir. The mechanisms include expansion of liquids and reservoir rock, natural energy from aquifers and gas caps, expansion of dissolved gas, and gravity drainage. Secondary recovery processes are often implemented by injecting water into the aquifer or injecting gas into the gas cap, to maintain the reservoir pressure. Recovery factors after primary drainage mechanisms and the implementation of secondary recovery techniques are generally moderate (Enick et al., 2012), hence there is a need for tertiary recovery techniques (Enhanced Oil Recovery, EOR) (Ahmadi et al., 2018). The latter aim to improve the recovery by acting on fluids and reservoir rock. Some of the most successful tertiary recovery techniques include water alternating gas injection, miscible CO_2 injection, polymer and surfactant injection, etc. (Afzali et al., 2018; Ahmadi et al., 2016; Dai et al., 2014; Ghriga et al., 2019; Vahdanikia et al., 2020; Xu, 1998). In addition to these three famous recovery stages of hydrocarbon reservoirs, other intervention strategies can be considered during the lifecycle of oil and gas reservoirs, mainly by infill drilling as well as the conversion of wells (e.g. producers into injectors, or vertical into horizontal) (Ding et al., 2020; Jesmani et al., 2020; Redouane et al., 2019).

The optimization of the recovery processes during the different recovery stages is crucial to optimize the techno-economic parameters such as Net Present Value (NPV) while taking into account the different constraints linked to production systems (pressure types such as Minimum Miscibility Pressure (MMP) and Bottom Hole Pressure (BHP) in miscible gas injection, as well as other production parameters such as water cut and Gas Oil Ratio (GOR), etc.) and the cost of the operation (cost of water injection, gas injection, well intervention operations, etc.) (Dai et al., 2014; Nait Amar et al., 2020c; Nait Amar and Zeraibi, 2019; You et al., 2020a, 2020b). Given the non-linearity of differential equations and thermodynamic models describing the different recovery processes, as well as the irregularity and heterogeneity of geometry (computational domain), the description and prediction of the evolution of key design parameters are commonly done numerically by using very powerful computing tools (Nait Amar et al., 2018a; Shahkarami et al., 2014a, 2014b). In this context, several commercial software such as Eclipse™ and CMG™ have been developed in the petroleum industry to allow a rigid optimization of the different tasks related to development strategies of reservoirs and production, while integrating advanced computing paradigms such as black oil, compositional, and streamline approaches. However, the optimization of a process described by a highly non-linear model with non-linear constraints and dependent on a significant number of parameters is very complex even using these advanced simulation tools (Panjalizadeh et al., 2015). Carrying out a direct simulation scenario with the latter for cases close to reality takes time and very efficient computing means (multiprocessors, parallel computing, etc.).

All the aforementioned technical constraints have led a great part of the petroleum community to investigate new alternatives which enable the same problems to be solved with considerable precision but with means that are not binding in terms of calculation time (Ertekin and Sun, 2019; Mohammadi and Ameli, 2019). Among these alternatives, Data-Driven Modeling (DDM) has gained increasing interest in the field of reservoir simulation. Approaches to DDM are generally statistics-based (or mathematics-based), e.g., the surface response method, and Machine Learning (ML) based. DDM may alternatively be known as proxy modeling while proxy model development englobes other approaches such as reduced-order modeling which mainly involve simplification of problems and are not purely data-driven.

The word proxy means to act on behalf of another. This definition has a projection on the technical or numerical sense of proxy models (also known as surrogate models). These are models built from data exploited from numerical simulations, capable of reproducing the simulator's responses with very high precision at a speed of execution that is in the order of a few seconds (Zubarev, 2009). These models have had vast use since the beginning of the 21st century in various areas. The use of proxy models has quickly been proposed in the field of reservoir engineering where there is a wide application of proxy models as substitutions for commercial software in various vital tasks such as well placement optimization (Hassani and Sarkheil, 2011; Sayyafzadeh, 2015a; Zarei et al., 2008), history matching (Sayyafzadeh, 2015b; Shahkarami et al., 2014b), and uncertainty studies (Mohaghegh et al., 2012a, 2006).

As proxy modeling can be regarded as a kind of pattern recognition and functionality identification, the model construction can be done with interpolation methods and Artificial Intelligence (AI) and ML methods. In this aspect, AI can be perceived as technology or tools that simulate the human brain and logic to perform analysis or any assigned task whereas ML denotes the application of computer algorithms to enable learning through data (Mohaghegh, 2018, 2017a, 2017b). Thus, ML is the subset of AI. The effectiveness of a proxy is very dependent on the robustness of the technique used for its elaboration (Na-udom and Rungrattanaubol, 2015; Zubarev, 2009). The robustness of an ML technique can touch upon various aspects, specifically the training procedure including the evolved relevant model parameters to improve the training and the considered mathematical operators (e.g., backpropagation process) in the calculation process. Artificial Neural Networks (ANNs), Support Vector Machines (SVM), kriging, and Response Surface Models (RSM) are among the widely applied techniques for building proxy models in the oil industry. In general, the first two approaches are ML-based whereas the last two are statistics-based. In this paper, our focus is on the ML-based proxy, also known as an intelligent or smart proxy[1]. It is worth mentioning that before proceeding to the building stage of the proxy, a primordial step consisting of generating a set of points or a database should be done properly. The judicious choice for sampling of the points will bring precision and generalization to the built model because the chosen sampling method tries to capture a wide variety of information about the inputs/responses of the simulators (Yeten et al., 2005). Design of Experiments (DoE) is the statistics branch assembled with proxy models through its methods (Crombecq, 2011; Forrester et al., 2008; Zubarev, 2009). Several works comparing different DoE methods have been published (Crombecq, 2011; Viana, 2016; Yeten et al., 2005). The main conclusion that can be retrieved from applying DoE in the building phase of proxy models is that space-filling techniques, such as Latin Hypercube Design (LHD), are one of the most efficient methods for building rigorous proxy paradigms. The details of the paradigm of intelligent proxy will be delineated later.

The optimization of different complex processes in the oil industry, such as EOR techniques, is a crucial step in reservoir management that significantly affects the efficiency and production strategy (Yazdanpanah and Hashemi, 2012). Several time-dependent parameters and the management procedure should be optimized in such projects (Yazdanpanah and Hashemi, 2012). Thus, traditionally the optimization methods evaluate hundreds or even thousands of potential scenarios to search for the optimal solution, using time-consuming numerical simulations. To deal with this issue which includes the significant calculation time and the considerable number of simulation runs, coupling metaheuristic algorithms with a powerful clustering-based proxy model is generally considered a better alternative for non-linear and multidimensional problems

[1] To avoid confusion, “intelligent proxy” (or intelligent model) and “smart proxy” models share the same definition in this paper.

(Onwunalu et al., 2008). Metaheuristic algorithms are population-based optimization techniques that consider a predefined criterion (fitness function) to distinguish between the performance of the individuals mimicking the scenarios of the problem. The gain of this kind of coupling is ensured by the exploitation of the advantages of the two approaches, namely the reduced calculation time of the proxy models, and the oriented and targeted runs to perform based on the fitness function of the metaheuristic algorithms. As discussed in (Onwunalu et al., 2008), a proxy model is employed to approximate the objective function values of different scenarios. When the estimated values exceed a certain threshold, the respective scenario will be chosen for simulation and optimization. Besides, it is worth mentioning that a smart proxy that is built using a significant number of numerical simulations can be used for dealing with uncertainties as the generated information is generally widespread and it involves an extensive number of interactions between the main parameters of the model for covering this kind of tasks.

Metaheuristic algorithms are the optimization algorithms we would like to emphasize in this work. Metaheuristics algorithms can be defined as mathematical frameworks with advanced searching mechanisms in the solution space (Gogna and Tayal, 2013; Wong and Ming, 2019; Yang et al., 2014). The advanced searching mechanisms of metaheuristic algorithms consist of the exploration and exploitation steps which involve specific operators that help the orientation of the optimization process towards regions of interest within the search space (Hemmati-Sarapardeh et al., 2020b). Exploration refers to inspecting the unexplored parts of the search space, while exploitation corresponds to the search of the neighborhood of the promising area (Tilahun, 2019). In general, these algorithms are derivative-free and nature-inspired. Examples of these algorithms include Genetic Algorithm (GA), Differential Evolution (DE), Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), Artificial Bee Colony (ABC), Firefly Algorithm (FA), Imperialist Competitive Algorithm (ICA), Simulated Annealing (SA), Gray Wolf Optimization (GWO), Cuckoo Optimization Algorithm (COA), etc. These algorithms have demonstrated their robustness in many areas of application, including prediction of stocks, image processing, bioinformatics, etc. (Gogna and Tayal, 2013).

In terms of reservoir simulation, metaheuristic algorithms have been extensively and successfully employed not only to train different types of proxies but also to solve optimization problems (coupled with either numerical models or proxies). For clearer perusal, implementation of metaheuristic algorithms in the establishment of ML-based proxies and resolution of optimization problem is conveniently termed the coupled ML-metaheuristic paradigm. Based on our studies (Nait Amar et al., 2021, 2020c; Ng et al., 2021a), the paradigm illustrated excellent results of implementation in developing ML-based proxy models where the metaheuristic algorithms were used for training. Additionally, optimization problems can be handled efficiently by applying the coupled ML-metaheuristic paradigm where this paradigm achieves optimum results within reasonable calculation time. Therefore, it is important to have a survey of how useful ML methods are to establish intelligent proxies when being solely employed or coupled with metaheuristic algorithms. Moreover, we opine that there is a necessity to provide this survey since there is not much available discussing these domains together.

This survey paper covers a wide range of research studies related to the application of ML techniques and the coupled ML-metaheuristic paradigm in intelligent proxy modeling. This work will contribute to the research and development related to various reservoir simulation applications mainly by shedding light on the smart schemes and intelligent methods based on ML and metaheuristic algorithms that were implemented for reducing the calculability efforts associated.

The rest of the paper is formulated as follows: Section 2 provides a brief discussion regarding some of the previous literature and reviews on the relevant topics. Section 3 demonstrates the general framework that can be employed to develop an intelligent model. Thereafter, Section 4 briefs several examples of the application of intelligent proxies and the coupled ML-metaheuristic paradigm in the context of reservoir simulation. Section 5 outlines the benefits and limitations of these paradigms as well as the associated challenges in the research domain before ending this survey paper with concluding remarks.

2. PREVIOUS WORKS

As briefly mentioned, Data-Driven Modeling (DDM) is considered another modeling approach aside from traditional physics-based modeling. The availability of a large database in petroleum engineering (Mohammadpoor and Torabi, 2020) has, to a certain extent, contributed to the prevalence of data-driven models as data is one of the main building blocks for the use of ML (Mohaghegh, 2022). Explicitly speaking, these data are applied to develop a model that can provide useful insights to petroleum engineers to do some engineering judgments. In the domain of reservoir engineering, DDM has provided a fast and efficient alternative for reservoir simulation (Mohaghegh, 2017a). More intriguingly, the coupling of the metaheuristic algorithms with ML-based data-driven models is another topic that is worth a discussion. To have a better outlook on the development of ML and metaheuristic algorithms[2] in the oil and gas industry, we will briefly discuss some relevant previous works and review papers.

2.1. Proxy Modeling

DDM is considered proxy modeling in the aspect of reservoir simulation. Using the proxy model as the substitute for Numerical Reservoir Simulation (NRS) has been applauded due to its quick computation and satisfactory accuracy of results (Mohaghegh, 2022; Nait Amar et al., 2021; Ng et al., 2022a). A simple illustration is displayed in Fig 1 to outline the relationship between proxy modeling and other terminologies that would be expounded on in the following subsections. The terminologies, such as Subsurface Data Analytics, Top-Down Modeling (TDM), and Smart Proxy Modeling (SPM), will be explained in detail in Section 2.3. ML is one of the approaches to proxy modeling. Zubarev (2009) provided a comparative analysis regarding the effectiveness of four different techniques of proxy modeling as the substitute for complete reservoir simulations. These methods included polynomial regression, multivariate kriging, thin-plate splines, and ANNs. He inferred that in history matching, the proxy models could perform reasonably well in a deterministic case but not in a probabilistic fashion. In the optimization of infill-drilling, the proxy models also illustrated reasonable performance, but the solutions were not optimal. Nevertheless, these models demonstrated excellent performance in terms of prediction of initial hydrocarbons in-place and oil recovery. In general, he stated that kriging models outperformed the others but induced the highest computational footprint. There was another constructive comment that the proxy modeling methods heavily relied upon the sophistication of the model, size of the design space, and quality of input data. This gives us a very well-established cognizance of the limitations or constraints that proxy modeling methods are subject to (Zubarev, 2009). He also opined that the option of proxy modeling methods was problem-dependent and quantifying the errors induced by proxy modeling techniques was needed for quality assurance.

[2] Based upon our survey of the literature, there are not many papers that solely discuss the coupling of metaheuristic algorithms with ML in the petroleum industry. Thus, in this survey paper, apart from explaining the use of ML, one of our discussions is intended to focus on how metaheuristic algorithms can be effectively implemented along with ML mostly in the context of reservoir simulation.

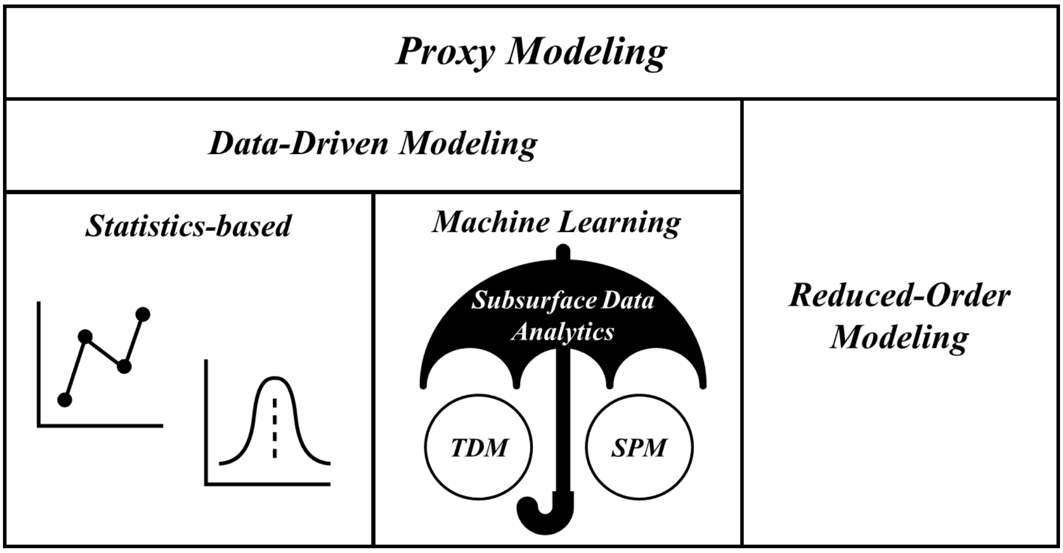


Fig 1 Schematic of the relationship between proxy modeling and other terminologies.

Moreover, Jaber et al. (2019a) conducted a detailed review of the application of proxy modeling in NRS. They summarized that there were two general approaches employed to develop proxy models, which included virtual intelligence and statistical method. Fundamentally, the proxy models were aimed at simplifying the complexity of the physical process regarding uncertain variables and assessing the responses rapidly with reasonable accuracy (Jaber et al., 2019a). The authors expounded that ANN, Fuzzy Logic, and GA were among the prevalent virtual intelligence methods used to build proxy models whereas RSM was the common statistical method in this context. In addition, they discussed several pieces of literature that illustrated the successful applications of virtual intelligence-based proxy models in assisted history matching and forecasting reservoir performance, and statistics-based proxy models in uncertainty analysis and prediction of reservoir response. They also outlined the proper step for validating and evaluating the quality of models. They further argued that virtual intelligence methods coupled with NRS were unable to simultaneously capture the effect of interactions among different uncertain variables. Hence, they opined that statistics-based proxies in general outperformed virtual intelligence-based proxies. More rivetingly, they shared the same opinion with Zubarev (2009) that understanding the use of a proxy was essential in choosing the right method, and evaluating the quality of proxies was highly recommended.

2.2. Implementation of ML

Apart from review papers about proxy modeling, several works expound on the general trend of the implementation of ML in the oil and gas industry. Li et al. (2020) provided an interesting insight into how rapidly the transition from digital oilfield to AI oilfield has taken place. Concerning this, they further outlined the pros and cons of different ML algorithms, including ANN, PSO, Fuzzy Logic, SVM, and GA. Thereafter, they discussed the efficient employment of AI in different aspects of the petroleum industry, e.g., history matching, dynamic prediction of production, optimization of a development plan, identification of oilfield development, detection of fracture, and EOR. In general, they inferred that compared to the other AI algorithms, ANN was the most prevalently used in the petroleum industry. Appropriate selection of the algorithm was also the solution to certain limitations of the algorithms. They further added that AI algorithms were too data-oriented and marginalized the physics of the process. More importantly, they pointed out that having the capacity to use and integrate big data of the oilfield with intelligent models at different phases was pivotal to ensuring the success of the AI oilfield.

Moreover, Ertekin and Sun (2019) conducted a painstaking status check on the implementation of AI in reservoir engineering. They presented different reservoir engineering-related research works, for instance, proxy modeling, AI-assisted history matching, and optimization of project design, which highlighted the robustness of the AI system. From this, they opined that the formulation of AI models could be divided into two distinct categories: forward and inverse-looking models. Additionally, data could be categorized into three groups, namely reservoir characteristics, project design parameters, and field responses. Perceiving the types of formulation and the associated data could provide a clearer understanding to the reservoir engineers in applying the AI approaches. Nonetheless, they arose the lack of astuteness of AI methods in completely replacing the traditional reservoir engineering models. Thus, they encouraged the hand-shaking protocol between the traditional modeling and the intelligent paradigm to fully exploit the respective advantages of each method and produce a more robust solution to reservoir engineering problems.

Furthermore, Balaji et al. (2018) evaluated the status and implementation of data-driven approaches, including ML, in the oil and gas industry. They first explained different data-driven techniques: linear regression, principal component analysis (PCA), decision tree, SVM, ANN, Fuzzy rule-based systems, GA, and Bayesian Belief Networks. Then, they showed how these methods were used in cases like subsurface characterization and petrophysics, drilling, production, reservoir studies and EOR, facilities, remediation and management, and pipelines. Pros and cons in tandem with the reasons for acceptance (as well as rejection) of these methods in the industry were also touched upon. More specifically, Alkinani et al. (2019) provided a review of the employment of ANN in the industry. They showed the basic steps in ANN modeling: collection and selection of input data, partitioning of data, normalization of data, and determination of the number of hidden layers and training algorithm. Also, they discussed the successful application of ANN in exploration, drilling, production, and reservoir engineering. In addition, Hanga and Kovalchuk (2019) thoroughly discussed the applications of ML and Multi-Agent Systems (MAS) in the petroleum industry. ML was proven to be effective in production, anomaly detection, and price detection while MAS was applied successfully in production, safety and maintenance, and supply chain management. They also stated how ML and MAS could be used interchangeably in various petroleum industry tasks and discussed the hybridization of both for better implementation. Apart from these, Otchere et al. (2021) did a detailed review of different pieces of literature to compare the application of ANN and SVM models in the forecasting of properties of petroleum reservoirs (mainly seismic and well log applications). They inferred that in the domain of reservoir characterization with limited data and in terms of coupling with other algorithms, SVM was found to outperform ANN.

2.3. Subsurface Data Analytics

Despite still having a lack of astuteness, the application of AI in petroleum engineering, especially for reservoir engineering, has gradually achieved enviable breakthroughs and maturity thanks to the contribution of the research group led by Dr. Shahab Mohaghegh. In this aspect, Mohaghegh (2011) explained the complete workflow that has been formulated to exploit the pattern recognition capabilities of AI in building an AI-based model that could act as a substitution for NRS. In this work, a constructive comment that was different from that of Li et al. (2020) regarding the use of physics was presented. He articulated that the use of physics was preserved through the generation of a spatio-temporal database. In simpler terms, it was denoted that the physics of the system was represented by the data. Hence, applying data with the help of AI to develop a model does not ignore physics. He further stated that the existing physical models (and statistical approaches) involved a lot of underlying assumptions which could have simplified the physics of the real problems. He has been consistently championing the utilization of data and AI because of his strong belief that the oil industry is heading toward the fourth paradigm of science, which is data-intensive science (Mohaghegh, 2020; Mohaghegh, 2011). Thus, he has systematized the whole idea of employing petroleum-related data in the establishment of models and coined it “Subsurface Data Analytics”. In general, the benefits of Subsurface Data Analytics over NRS, including circumvention of preconceived notions, biases, and simplifications of problems, have been highlighted.

Mohaghegh (2020) expounded a deep concern regarding the hybrid models (combination of physics-based and AI-based approaches) and opined that hybrid modeling was the conventional statistical approach. The reasons for building hybrid models were assumed the lack of ability in developing good models by only applying ML techniques, employing it as a marketing tool, lack of ability in explaining the results produced by the ML-based models, and lack of ability in responding to the challenges imposed by the conventionalists in the industry.

Under the umbrella of Subsurface Analytics, there are two main classes of modeling, which are Smart Proxy Modeling (SPM) and Top-Down Modeling (TDM). According to Mohaghegh (2018, 2017a, 2017b), the formulations of both TDM and SPM share the same fundamental idea and methodology. Both models are defined as an ensemble of Neuro-Fuzzy systems that can learn and recognize the hidden pattern of the data provided. The only subtle difference is the source of the data used. For SPM, the data come from the spatio-temporal database generated by NRS whereas the spatio-temporal database for TDM originates either from the field data or the combination of field and simulation data. Regarding the functionalities of these two types of proxies, the smart proxy model is mainly implemented to reduce the computational effort induced by NRS while producing outputs within a satisfactory level of accuracy (Mohaghegh, 2018, 2022). This rapid and accurate assessment can help reservoir engineers to elude wasting extra time in making some reservoir management-related decisions. Besides that, the relevant details of TDM have been outlined in this literature (Mohaghegh, 2017a). It is a completely different method of modeling a subsurface as compared to NRS using a bottom-up approach. In general, TDM is applied to develop a model that can better decipher the behavior of the reservoir system. Both SPM and TDM are useful in different reservoir engineering tasks, including history matching (He et al., 2016; Shahkarami et al., 2018), CO_2 storage and sequestration (Mohaghegh, 2018), CO_2-EOR (Shahkarami and Mohaghegh, 2020; Vida et al., 2019), and shale analytics (Mohaghegh, 2013). There is also an associated challenge with both TDM and SPM in which the curse of dimensionality will happen as the size of the spatio-temporal database increases. In this case, Mohaghegh (2018, 2017a, 2017b) initiated the use of fuzzy pattern recognition to determine the degree of influence of each possible parameter on the output in terms of Key Performance Index (KPI). The ranked KPIs aid in selecting the input variables. Using fuzzy logic is preferred when calculating the KPIs of input variables because it can model uncertainties associated with vagueness or lack of information as discussed in these references (Mohaghegh, 2018, 2017a, 2017b; Ross, 2010).

The generation of massive data, fathomed as “Big Data”, in the upstream and downstream petroleum industry has also played an integral part in the emerging trend of the use of ML in the industry. In this case, Mohammadpoor and Torabi (2020) illustrated a comprehensive review of how Big Data analytics has been effectively utilized in the industry. They expounded on six characteristics of Big Data that included volume, velocity, variety, veracity, value, and complexity. They outlined the general methodology of Big Data and explained the tools that could be used to perform Big Data analytics. They also presented different examples to demonstrate how it was implemented in different aspects of upstream, such as exploration, drilling, reservoir engineering, and production engineering. Examples of downstream were also provided, e.g., refining, oil and gas transportation, and health and safety execution. Besides that, Temizel et al. (2016) explained the general steps involved in Data Mining and the development of data-driven models through the illustration of a synthetic case. Apart from briefly explaining the use of statistics-based and ML-based methods in Data Mining, they also conveyed the fundamental thought of how data could be useful in terms of modeling if being systematically used. More intriguingly, Ani et al. (2016) discussed the importance of applying uncertainty analysis (probabilistic approaches) in reservoir modeling compared with the deterministic approach. In this context, they added that the use of ML would have a significantly positive impact on the future trend of uncertainty analysis.

2.4. Application of Metaheuristic Algorithms

Based on our investigation, the literature comprehensively reviewed the successful use of metaheuristic algorithms in different domains. However, there are only a handful of studies that examined their application along with ML, especially in the field of petroleum engineering. The metaheuristic algorithms discussed in this paper are mainly nature-inspired. We opine that these algorithms are robust in terms of implementation. They are not only widely used in optimizing the hyperparameters of the intelligent models (Hemmati-Sarapardeh et al., 2020a; Nait Amar et al., 2018b; Nait Amar and Zeraibi, 2018; Ng et al., 2022b, 2021c), but also in solving petroleum engineering-related optimization problems (Nait Amar et al., 2021, 2018a; Ng et al., 2021b; Wang et al., 2021). Hemmati-Sarapardeh et al. (2020b) included an extensive explanation of the mechanism of different metaheuristic algorithms, such as GA, PSO, ACO, ABC, FA, and GWO. They also illustrated how these algorithms could be coupled with different intelligent models and employed in different domains like reservoir and production engineering, drilling engineering, and exploration. Moreover, Plaksina (2019) performed a similar review but with more emphasis on evolutionary computation, swarm intelligence, fuzzy logic, different types of ML, and ANN. She included a lot of petroleum-related applications concerning the abovementioned areas to illustrate the robustness of AI approaches. Also, Rahmanifard and Plaksina (2019) reviewed and explained different optimization approaches, such as GA, DE, and PSO in tandem with ANN and fuzzy logic. They also provided some discussions to outline the applications of these methods in the petroleum industry.

3. PARADIGM OF INTELLIGENT PROXY DEVELOPMENT

In this section, we will brief the general framework used in establishing intelligent proxy models in the context of reservoir simulation. This framework is a product of assimilating different workflows proposed in several pieces of literature (Hemmati-Sarapardeh et al., 2020b; Mohaghegh, 2017a; Russell and Norvig, 2010). In this aspect, when ML methods are implemented to perform proxy modeling, it can be termed as either "smart" or "intelligent". The word "smart" or "intelligent" indicates the capability of the model to learn and decipher the hidden pattern or relationship between the input and output data provided using the ML methods. Metaheuristic algorithms can act as training algorithms to help the models learn better. Their robustness is demonstrated as they can conveniently be coupled with the built intelligent models to solve optimization problems. As mentioned earlier, data act as the most essential element required to build the intelligent proxy model. Hence, it is of paramount importance that the data provided to the proxy correctly capture and represent the physics of the system being modeled. Besides that, we need to understand that the intelligent proxy is never a one-size-fits-all model. The fundamental paradigm of building an intelligent proxy is summarized in Fig 2.

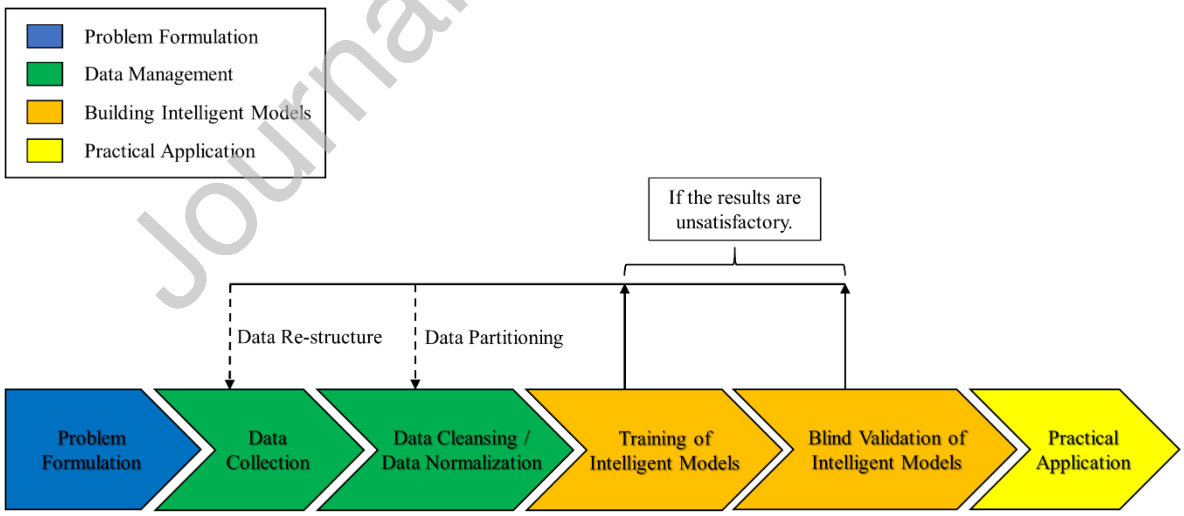


Fig 2 Paradigm of Intelligent Proxy Development.

The first step of the paradigm is to identify the purpose of the proxy and carefully ***formulate the problem***. This is important because it provides a clear idea regarding the type of database that needs to be generated or extracted. Having defined the optimization problem clearly, the reservoir engineer would have a better perception of the data needed to develop the corresponding proxy model. It is also vital to emphasize that the number of proxy models required depends upon the complexity of the formulated problem. The important takeaway of this step is that one should be cognizant of the problem to be solved, define it clearly, and ensure the proper variables or parameters needed to build the proxy. Besides that, selecting the appropriate AI methods is another consideration in this step. Such appropriateness can be determined by the capability of the selected method to mathematize the relevant engineering problem as a functional relationship.

There are two main categories of input variables for reservoir simulation, namely static and dynamic input data. Examples of static data include porosity, permeability, and thickness of the formation layer. Dynamic data consist of production rate, well bottom hole pressure, and saturation. It is important to understand that if a dynamic parameter is considered one of the input variables, one might require to develop a proxy that can forecast this dynamic variable. Thereafter, the predicted dynamic parameter should be fed into the initial proxy to reduce the dependency on the use of NRS. This type of proxy design is termed "cascading design" (Mohaghegh, 2017a). Nonetheless, one should be aware of the possibility of accumulation of prediction error when the "cascading design" is employed. Therefore, these points of discussion ought to be pondered ahead during the phase of problem formulation to ensure a smooth process of proxy development in later stages.

Then, as we proceed to ***Data Management***, we need to understand the types of data that should be obtained and identify the sources of data to retrieve the database (NRS, field measurements, or both). In this paper, our discussion concentrates on the use of data generated by NRS. To generate the database from NRS, we need to design several scenarios of simulation runs. Thus, we implement a sampling strategy to extract several samples (of for example rates) within the predefined operational range and define them as simulation scenarios. Each scenario is equivalent to one simulation run. Based on our survey, there is no specified number of runs required to create the database. Theoretically, the higher the number of simulation scenarios, the higher the chances that the solution space of the optimization problem is covered. Nonetheless, this will cause the curse of dimensionality. So, the choice of sampling strategy plays a vital role in ensuring the success of proxy modeling. Examples of renowned sampling methods include Latin Hypercube sampling (McKay et al., 1979), Halton sequence (Halton, 1960), Sobol sequence (Sobol', 1967), and Hammersley sequence (Hammersley and Handscomb, 1964). The selection of input data (also termed feature selection) is another consideration in this step. During problem formulation, we would have known the output data that our developed proxy models can generate. It is important to identify the input variables with a larger degree of influence on the output. In terms of NRS, the selection of useful input variables is indeed essential because including too many of them will induce the curse of dimensionality. There are three approaches to this selection, namely empirical selection, statistical methods, and AI-based methods. The first selection relies upon common knowledge of reservoir engineering. Moreover, several statistical metrics, e.g., percentile of the highest score, k highest score, and chi-squared test are employed to select the useful input parameters. AI-based approaches such as fuzzy pattern recognition have shown successful and robust applications in choosing the input variables (Mohaghegh, 2018, 2017a, 2017b; Mohaghegh, 2011). According to our investigation, any of these three methods can contribute to the successful development of proxy models. However, Mohaghegh (2018, 2017a, 2017b) opined that fuzzy pattern recognition outperformed the statistics-based approaches in this context.

Before feeding the database into the model, data cleansing occasionally might be needed to remove any noisy data or outliers which can affect the learning of the models. This is normally done on real field data. For NRS, data cleansing is not needed. Data normalization is another highly recommended step before proceeding to the training of intelligent models, where the values of the database will be rescaled within a smaller range of values, generally either [0, 1] or [-1, 1]. Our survey based upon numerous papers (Hemmati-Sarapardeh et al., 2020b; Nait Amar et al., 2019, 2018b) confirms that data normalization is very common to ensure that the intelligent models can capture the pattern induced by the database. In this case, we highly suggest conducting "categorical normalization". For instance, when there are several columns of input data indicating the same category of data such as porosity, the maximum and minimum values of the datapoint should be chosen from the same category for normalization. After the completion of this phase, the database is deemed ready to be implemented for training the intelligent proxy models.

In the step of ***building intelligent models***, the fundamental idea is to enable the intelligent models[3] to learn and capture the physics of the system. Concerning this, it is important to perceive the definitions of model parameters and model hyperparameters (Yang and Shami, 2020). Model parameters refer to the ones that can be initialized and updated through the training process (viz. weights and biases for ANN). Model hyperparameters must be initialized before training and are related to the architecture of ANN, for instance, the number of hidden layers and nodes, learning rate, and dropout rate (Yang and Shami, 2020). Searching for the optimal model hyperparameters, alternatively known as hyperparameter optimization, can be performed to ensure better

[3] An example of intelligent models of interest here is ANN. However, the methodology also applies to other ML-based models.

learning ability of an intelligent model during training. The algorithm selected to perform such optimizations will iteratively tune the model parameters and model hyperparameters to minimize a predefined loss function until a stopping criterion is met. Examples of the loss function can be the Mean Squared Error (MSE), Mean Absolute Error (MAE), Average Percent Relative Error (APRE), and Average Absolute Percent Relative Error (AAPRE). In general, the algorithms used are categorized into two groups, including derivative-based and derivative-free. Examples of derivative-based algorithms include stochastic gradient descent, scaled conjugate gradient, Levenberg Marquardt, and Adaptive Moment Estimation whereas derivative-free algorithms are mainly nature-inspired, such as Genetic Algorithms and Particle Swarm Optimization. Also, combining both can be another option (Nait Amar et al., 2018c, 2018b).

The database needs to be partitioned into for instance three different sets, namely training, validation, and testing[4]. Albeit there is no rule for partitioning ratio, most of the literature (He et al., 2016; Mohaghegh, 2017a; Ng et al., 2022b, 2021c; Shahkarami and Mohaghegh, 2020) used either the ratio of 7:1.5:1.5 or 8:1:1. After the partitioning is done, the training data should be fed into the intelligent model to undergo the training phase. During this phase, for every iteration, the performance metrics of validation are evaluated to check if the overfitting issue occurs. Regarding this, we can infer that the overfitting issue is eluded if decreasing trends of loss function for both training and validation data are observed. If such a trend is not noted, training needs to be repeated. Refer to Shahkarami and Mohaghegh (2020) for the pertinent details. Nonetheless, before repeating the training, the dataset can be re-partitioned to evaluate if better training results can be yielded. However, such re-partitioning is regarded as bad practice by Russell and Norvig (2010). Thus, the whole training process can be performed by either adding new data points or using a completely new set of data (termed data re-structure)[5]. When the overfitting issue is assured to be prevented, we can deduce that the trained intelligent model has passed the first stage of quality assessment.

Training and validation performance can be assessed using the metrics used as the loss function in addition to the coefficient of determination, R^2. The use of APRE and AAPRE needs attention, especially during the establishment of a proxy model that predicts the water production rate. This is because the water production rate is zero just before the water breakthrough, given the initial water saturation is equal to the immobile water

[4]Alternatively, Mohaghegh (2017a) uses the terms calibration and validation datasets for validation and testing dataset, respectively.

[5] It relies upon the preference if re-partitioning of data should be attempted. In this work, our objective is to outline a general workflow that helps the readers to apply the approaches.

saturation. Thus, it can be cumbersome to implement APRE and AAPRE to evaluate the performance of proxy models before the water breakthrough. In this case, the testing data is fed into the model to evaluate the testing phase performance. This phase is to ensure that the trained model portrays a good level of predictability before being blind-validated, which is the last stage of quality evaluation. In blind validation, it is important to note that the blind data should not have been part of the training, validation, and testing data. Additionally, it is highly recommended to ensure that the blind validation dataset falls within the range of the previously generated database. This is because according to some literature (Barnard and Wessels, 1992; Haley and Soloway, 2003; Xu et al., 2020), intelligent models generally perform well in interpolation but not in extrapolation. If the result of blind validation is excellent, then it denotes that the model has good predictability to serve its purpose and is ready for practical application. Nevertheless, if the blind validation results are not satisfactory, data re-partitioning or data re-structuring can be considered. Generally, these three phases of the quality assessment provide insights to confirm that the model can serve its objective.

For the case of hyperparameter optimization, based on our study (Ng et al., 2021a), using the weighted sum of the training, validation, and testing errors are recommended. The respective weighting factors can be treated as additional parameters to be optimized. Also, one needs to understand that performing such optimization tasks will require additional time, proportional to the size of the database (Shahkarami and Mohaghegh, 2020). Therefore, there is a trade-off to consider when it comes to conducting the optimization. It is also important to know that the models can be divided into static and dynamic types. Static proxy models are usually built to predict specific variables over a whole period. For instance, a model that forecasts the NPV of a certain production period considering several input variables. This type of proxy is not robust in terms of application despite the ability to speed up the computation. Dynamic proxy models are established to forecast variables at certain timesteps. Albeit building them can be more laborious than static proxies, dynamic proxies offer higher flexibility in terms of application, including prediction of specified output and optimization (Nait Amar et al., 2018a). It is, therefore, necessary to highlight the distinction between these two types of proxies that helps one to have a better perception at the beginning of proxy modeling. Some examples of ***practical applications*** will be discussed in Section 4.

4. SURVEY OF APPLICATIONS

Applications of ML and coupled ML-metaheuristic paradigm in different domains of reservoir engineering, mainly in the areas that implement reservoir simulation, will be discussed here. Fig 3 illustrates the examples of domains that are surveyed in this section. Due to the limited use of coupled ML-metaheuristic paradigm, emphasis is on ML in several application examples. A few interesting works (discussing only the use of metaheuristic algorithms or their applications with other variants of proxy models, e.g., reduced-order modeling) have also been included in this section. The summary of the collected literature is demonstrated at the end of each subsection along with the methods used as well as the assumptions and limitations discussed in each work. Refer correspondingly to Table 1 to 8 for the summary of the literature on each subsection.

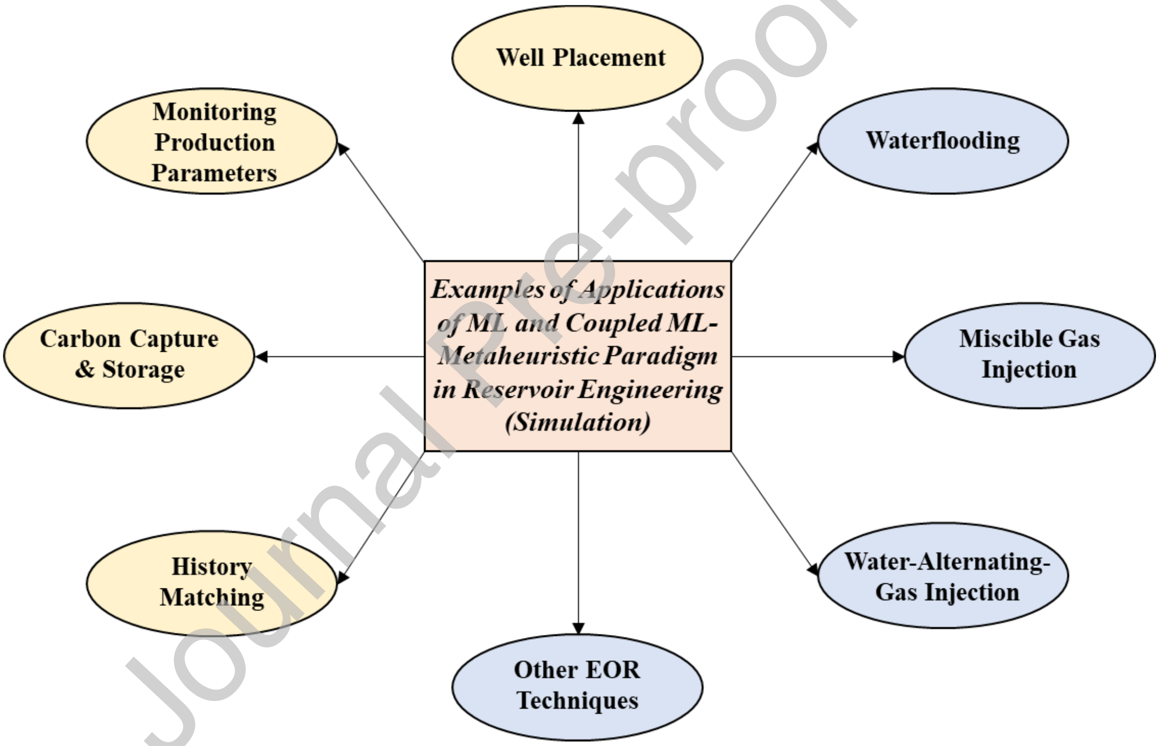


Fig 3 Examples of Application of ML and Coupled ML-Metaheuristic Paradigm.

4.1. Well Placement

Optimizing well placement is one of the most challenging tasks in field development planning. This is because multiple scenarios of NRS need to be run to determine the best location to place the wells. The computational efforts will increase when the geological uncertainty of the reservoir being modeled is considered for better decision-making. The optimization task can be cost-effective if the computational time can be shortened.

Several pieces of the literature suggest the application of ML approaches as the potential solution. Additionally, the coupling of the simulation models or the respective proxy models (built using ML) with the metaheuristic algorithms has shown some promising results.

Nwachukwu et al. (2018a) performed a handful of NRS to generate training data and implemented the Extreme Gradient Boost (XGBoost) approach to establish a model that could provide a fast forecast of the responses of a reservoir based upon the locations of injectors. In addition, they employed the Fast-Marching Method (FMM) to introduce the well-to-well connectivity to the model and this enhanced the results significantly. The methodology was used in the cases of waterflooding and CO_2 flooding. Thereafter, Nwachukwu et al. (2018b) extended this ML approach to optimize the location of wells and the parameters of WAG injection by coupling the model with a novel optimization algorithm, namely Mesh Adaptive Direct Search (MADS). Xiong and Lee (2020) applied the ANN modeling to build a model to estimate the production of fluids based on reservoir heterogeneity and well locations. Then, they used this model to determine the optimal location of injectors in the case of waterflooding. Chu et al. (2020) discussed the use of Convolutional Neural Network (CNN) to develop three different models, single-, dual-, and multi-modal CNNs, in the optimization of infill well locations. They also compared these models with a Feedforward Neural Network (FNN). Jang et al. (2018) proposed the sequential employment of ANNs to determine the optimal well location in a coalbed methane (CBM) reservoir. They inferred that the sequential ANNs computationally outperformed the direct use of PSO algorithms in the same optimization problem.

Sayyafzadeh (2015a) presented a self-adaptive surrogate-assisted evolutionary algorithm to determine the optimal location of wells. This algorithm was established by partially or fully replacing the original fitness function (OFF) with the approximate function (AF), which was represented by ANN. Then, two surrogates were used to stochastically decide whether OFF or AF would be applied. This methodology performed well on GA for the problem of optimizing well placements. Redouane et al. (2019) successfully suggested a newly enhanced intelligent framework that involved GA, design of sampling, and proxy model to achieve optimization of well placement in a fractured unconventional reservoir. Busby et al. (2017) illustrated the use of K-medoid algorithm to select the features to run the corresponding simulation and applied the data to train the ML algorithms such as neural networks, gradient boosting, and random forest. This data analytics workflow was successfully applied to a synthetic green field and showed that the location of wells could be optimized under uncertainty. In the work of Mousavi et al. (2020), XGBoost was shown to outperform the central composite design (CCD) method in determining the best location of wells under different reservoir scenarios. In other words, XGBoost could converge to the optimal solution compared with CCD. Kristoffersen et al. (2020) discussed how the methodology of Automatic Well Planner (AWP) could be employed in a specific type of neural network, known as Neuro Evolution of Augmenting Topologies (NEAT). By coupling the neural network model with a derivative-free algorithm, namely Asynchronous Parallel Pattern Search (APPS), the well placement decision was made optimally.

The potential implementation of metaheuristic algorithms is not limited to the above-mentioned pieces of literature. Pouladi et al. (2017) suggested the use of Fast Marching Method (FMM) to develop a proxy model and coupled the proxy with PSO to optimize multiple production well placements. Hassani et al. (2011) developed three different proxy models, such as quadratic model, multiplicative model, and radial basis function of a fractured reservoir in the west of Iran, and coupled the proxies with GA to optimize the horizontal well placement. Morales et al. (2010) also performed horizontal well placement optimization in gas condensate reservoirs with a modified genetic algorithm. They extended the use of the algorithm by considering a similar optimization problem under geological uncertainties (Morales et al., 2011). The literature on Well Placement is summarized in Table 1.

Table 1 Summary of Literature in the Domain of Well Placement.

Literature	Methods	Remarks	Assumptions / Limitations
Nwachukwu et al. (2018a)	XGBoost	With different well configurations, ML models with connectivities were built to predict different responses, viz. total profit, cumulative oil/gas production, or net CO_2 stored with less computational effort.	• Augmentation of predictor variables due to the sophistication of response surface. • The proposed methodology requires further verification in terms of optimization.
Nwachukwu et al. (2018b)	XGBoost / MADS Algorithm	An extended work of Nwachukwu et al. (2018a) in which ML models were made to offer reservoir responses corresponding to well locations and control during WAG under geological uncertainty. MADS was then used for joint optimization.	• Augmentation of predictor variables due to the sophistication of response surface. • Case-sensitive application. • The proposed methodology was implemented on a synthetic case.
Xiong and Lee (2020)	ANN	ML models were built to forecast fluid production as a function of heterogeneity and the location of the injector with an improvement of prediction	• Updating of models is needed when new data is available in the case of actual field data. • Verification of the suggested methodology

		accuracy by using data from injectors and producers. The selection of optimized injection well placement was done with the aid of P90 and P50.	with other strategies is required.
Chu et al. (2020)	FNN/ CNN	Multi-modal CNN outperformed FNN in terms of finding the optimal infill well placement.	• The study was only focused on a single vertical infill well. • Dynamic properties utilized as input data were obtained at the time of infill drilling. • The exponential increase of the size of search space if horizontal drilling is considered.
Jang et al. (2018)	Sequential ANN/ PSO	Sequential ANN modeling was implemented to refine the model developed. It outperformed the coupled paradigm between the simulator and PSO in terms of the number of simulation runs.	• The study was only conducted on a coalbed methane (CBM) reservoir. • The performance of the sequential ANN is influenced by its parameters which are meant to be tuned.
Sayyafzadeh (2015a)	FNN/ GA	A self-adaptive surrogate-assisted evolutionary algorithm was introduced to solve the well placement optimization problem with an improvement in accuracy.	• The study was only conducted on the PUNQ-3S reservoir. • For this optimization problem, infill wells were located at an equal distance. • The methodology is yet subject to verification of real-life cases.
Redouane et al. (2019)	Gaussian Process/ GA	An adaptive surrogate reservoir modeling was displayed to manage well placement problems in a real-life fractured reservoir model.	• Fixed cost of drilling and location independent costs. • The methodology is yet to be tested for other field development problems. • Formulation of different constraints, including well length, inter-well distance, reservoir bound, and well orientation.
Busby et al. (2017)	Neural Networks, Random Forests, Gradient Boosting / K-medoid algorithms	Data analytics workflow was shown to determine the locations of wells for a green field.	• Limited interaction between the wells to reduce the number of combinations. • Limited application to real field cases.
Mousavi et al. (2020)	XGBoost	An ML model was established to predict the NPV of a well placement problem through different scenarios for optimization purposes.	• Operational constraints of field development strategies were considered for the reservoir scenarios. • Only three scenarios were implemented.

Kristoffersen et al. (2020)	ANN/ APPS, PSO	Automatic Well Planner (AWP) was developed to increase the efficiency of well placement optimization under geological uncertainty.	• Formulation of constraints, such as length, dog-leg severity, and deviation of well. • The spatial distribution of self-selected properties is assumed to be defined within the reservoir model as property maps. • Wells were drilled at the beginning and at the same cost. • Fixed prices.
Pouladi et al. (2017)	FMM/ PSO **Although FMM is not considered ML, this paper showed the potential implementation of PSO in terms of coupling with any type of proxy model, which is worth reading.*	FMM-based proxy models were coupled with PSO to resolve well placement optimization with a very good computational efficiency.	• Darcy flux is assumed negligible for the volumetric pressure drop estimation by FMM. • It appeared to be impractical to illustrate the final pressure map for problems with more than one well.
Hassani et al. (2011)	Quadratic, multiplicative, and Radial basis function / GA	A proxy modeling approach was employed to enable the optimization of horizontal well placement to be handled more quickly.	• Models (to estimate cumulative oil) are assumed to be a function of the location, direction, and length of a new horizontal well. • The proposed methodology is yet to be tested for multiple geological realizations.
Morales et al. (2010)	GA	A modified GA was employed to optimize a horizontal well placement in a Gas Condensate reservoir. The Minimal Variation (MiniVar) was modified in this case.	• The wellbore was set as eight grids in length. • Deterministic approach. • Published data of the field is limited.
Morales et al. (2011)	GA	A slight extension of Morales et al. (2010) in which geological uncertainty was considered.	• Published data of the field is limited. • Assumption of the probability of success and weights assignments to each realization.

4.2. Monitoring Production Parameters

In reservoir engineering, hydrocarbon and water productions play a pivotal role in determining the economic feasibility of a field development project. In this context, hydrocarbon production parameters such as oil and gas production rates must be monitored carefully to ensure substantial financial returns for the plan. Water production needs to be monitored to avoid unnecessary handling costs. Therefore, it is essential to develop a model that can monitor and predict these production parameters. However, solely applying the conventional physical and mathematical approaches to build the model is indeed challenging. The reason is that the complexity of the system has been simplified by some assumptions to justify the validity of the physical model. This is where ML methods can be applied to elude the use of these simplifications. Some literature have illustrated the successful applications of ML in monitoring and forecasting production parameters. Some applications also highlighted the development of the models by coupling the ML methods with metaheuristic algorithms.

One of the traditional approaches in production forecasting is decline curve analysis (DCA). However, Mohaghegh (2017a) explained that DCA might be insensitive to some changes in operational conditions during implementation. Therefore, ML has been preferred as an alternative for monitoring and production forecast. Sun et al. (2018) implemented the Long Short-Term Memory (LSTM) algorithm to develop a data-driven model to predict the production rate of multiple wells by only employing the production history and tubing head pressure as the input variables. Thereafter, they compared the yield of the data-driven model with three DCA models, which are Duong model, Power Law Exponential Decline (PLE), and Stretched Exponential Decline (SEPD). The comparison illustrated that the LSTM model produced the production forecast with higher accuracy. Alkhalaf et al. (2019) successfully demonstrated the application of ANN in well production forecasting by feeding the real-time data into the model. They also performed the grid search method to optimize the architecture and hyper-parameters in modeling the ANN. Masini et al. (2019) showed the successful use of XGBoost to build a data-driven model to replace DCA. In their work, clustering techniques such as Random Forest and Density-based clustering had been used to cluster the data points with close operational conditions before training the model to conduct DCA. More intriguingly, Omrani et al. (2019) applied the hybrid approach, which was the combination of a physical model (nodal analysis) and ANN, to predict well production. They inferred that the hybrid approach performed better for long-term production forecasts (production of several years).

The use of ML approaches is extended to other domains of production engineering. Khan et al. (2019) employed ANN, SVM, and Artificial Neuro-Fuzzy Inference Systems (ANFIS) to estimate the oil rate in the artificial gas lift wells. They observed that ANN yielded much better results compared with SVM, ANFIS, and other empirical models. Furthermore, ML methods can be implemented to forecast measurements obtained from virtual flow metering and permanent downhole gauges. Bikmukhametov and Jäschke (2020) examined different approaches to hybridizing ML with first principles models of process engineering to successfully predict the volumetric flow from Virtual Flow Meter. Additionally, Tian and Horne (2017) utilized the information from permanent downhole gauges to develop a data-driven model to forecast reservoir performance via the application of recurrent neural network (RNN). Alakeely and Horne (2020) showed the potential of RNN by employing it to simulate the behavior of reservoir model. CNN was also implemented and demonstrated good results. Yang et al. (2019) illustrated a novel method in which advanced mud gas data was used to develop an ML model to estimate GOR effectively. The model comprised a combination of different techniques such as Gaussian Process, Universal Kriging, Random Forest, K-Means Regressor, and Elastic Net-regularized linear regression model. Chen et al. (2019) proved the excellent integration of ANN modeling with conventional reservoir analog studies to conduct recovery forecasts. The unsupervised ML method, autoencoders (AE), was shown useful by Alatrach et al. (2020) in predicting well production events. In this work, a 6-layered AE-NN model demonstrated positive results and could detect the deviation from the expected behavior of a well.

There are also some literature discussing the use of ML techniques in unconventional resources. Rahmanifard et al. (2020) performed a design of experiment to develop an ANN model that accurately approximated the well production in Montney Formation, a shale gas formation. Cross et al. (2020) successfully built a decision tree-based ML model to forecast the water production of a well in Williston Basin. Another ML technique, which was the partial least square (PLS) algorithm, was employed by Al-Alwani et al. (2019) to predict the production performance in Marcellus shale based on parameters obtained from stimulation and completion. ANN modeling was employed by Cao et al. (2016) to develop data-driven models for two different scenarios, namely prediction of future production of an existing well and production forecast of a new well. They demonstrated that by incorporating the geological features, the production forecast of new wells produced excellent results. Amaechi et al. (2019) applied ANN and Generalized Linear Model (GLM) to estimate the initial gas production rate from tight gas reservoirs in Ordos basin. They implemented Garson Algorithm in ANN and Variable Importance in GLM to identify the KPI of each feature used in the development of models. The robustness of ML techniques in unconventional resources has been further validated when Urban-Rascon and Aguilera (2020) used ML to build models to achieve optimization in stimulated reservoir volume (SRV) characterization, discretization of fracture systems, and production prediction. In their work (Urban-Rascon and Aguilera, 2020), a self-organizing map (SOM) was utilized to map the hydraulic fracturing stages with microseismic data. Chaikine and Gates (2021) used a hybrid model of convolution-recurrent neural network (c-RNN) to forecast the production from multi-stage horizontal well whereas Hassan et al. (2019) employed ANN to estimate the well productivity of fishbone wells. This literature highlighted the wide applicability of ML in production engineering. These ML methods can also be coupled with metaheuristic algorithms to be more fruitful. Han and Bian (2018) developed a hybrid model of SVM and PSO to estimate the oil recovery factor in a tight reservoir. Panja et al. (2018) applied PSO to optimize the hyperparameters of SVM and the weights and biases of ANN, which were used to predict the production from shale plays including Eagle Ford, Niobrara, and Bakken in United States. Refer to Table 2 for the summary of the literature on Monitoring Production Parameters.

Table 2 Summary of Literature in the Domain of Monitoring Production Parameters.

Literature	Methods	Remarks	Assumptions / Limitations
Sun et al. (2018)	RNN-LSTM	Comparing the production forecast of multiple wells between DCA and RNN-LSTM.	• Assumption of constant tubing head pressure. • Assumption of initial production for a few years in a few wells.
Alkhalaf et al. (2019)	ANN	Using ANN to predict the flow rates.	• The process of retraining is limited to a predefined threshold or every ten new real-time measurements.
Masini et al. (2019)	Random Forest, XGBoost	Demonstrating automated DCA by using ML methods.	• Requiring the specification of parameters for every new data set. • Limitation of data set: only choke data available.
Omrani et al. (2019)	ANN	Hybridizing the first principle model and ANN to predict the short-, mid-, and long-term production.	• Limited training sets. • Assumption of production and operational conditions.
Khan et al. (2019)	ANFIS, ANN, SVM	Applying ML to predict the oil rate in the artificial gas lift.	• Limitation of the number of epochs to 400. • Limited data sets.
Bikmukhametov and Jäschke (2020)	Gradient Boosting, ANN, LSTM	Combining the ML models with the physics of process engineering to forecast the multiphase flow rates.	• Simplification of the first principle models. • Assumption of steady-state flow and negligible effect of the acoustic wave.

Tian and Horne (2017)	RNN	Employing RNN for the data analysis of permanent downhole gauge.	• Assumption of model parameterization. • Models were for case-specific applications.
Alakeely and Horne (2020)	CNN, RNN	Using CNN and RNN to simulate the reservoir responses.	• Limited amount of data. • Models were for case-specific applications.
Yang et al. (2019)	Gaussian Process, Kriging, Random Forest, K-Mean, Elastic Net	Implementing machine learning to predict gas oil ratio based on advanced mud gas data.	• Limited gas input data. • Limited data collection. • Limited application to formation-wise model.
Chen et al. (2019)	ANN	Forecasting the reservoir recovery by using ANN based on the analog study.	• The number of ANN hidden layers was limited to 3. • Assumption of the development of reservoir database through a large number of well patterns.
Alatrach et al. (2020)	Autoencoders	Predicting the event of well production by using autoencoders.	• Data from limited wells. • Occurrence of false positive prediction (training was conducted on some missed events of production).
Rahmanifard et al. (2020)	ANN	Forecasting the well performance in Montney Formation.	• Models were for case-specific applications.
Cross et al. (2020)	Decision tree-based model	Prediction of water, gas, and oil production at a timestep of 30 days for the first two years in the Williston Basin.	• Lacking information about water-related geology features for more robust modeling. • Models were for case-specific applications.
Al-Alwani et al. (2019)	Partial Least Squares (PLS)	Estimating the performance of production in Marcellus Shale from stimulation and completion parameters.	• Limitations in the database, including percentage parameters exceeding 100%. • Limited use of P10, P50, and P90 production forecast.
Cao et al. (2016)	ANN	Production forecast using ML in unconventional reservoirs.	• Data consisting of operational constraints. • Production history of the well was needed as a starting point in the case of ANN.
Amaechi et al. (2019)	ANN, GLM	Estimating the initial gas production rate from tight reservoirs.	• Models were for case-specific applications.
Urban-Rascon and Aguilera (2020)	SOM	Production prediction in low permeability reservoirs.	• Assumed that earthquake showing self-similar behavior in fracture scaling.

			• Models were for case-specific applications.
Chaikine and Gates (2021)	c-RNN	Using c-RNN to forecast the production from multi-stage hydraulically fractured horizontal wells.	• Limiting the number of variables used. • Limited sample sizes.
Hassan et al. (2019)	ANN, Fuzzy Logic, RBF-NN	Well productivity forecast from fishbone wells using ML methods.	• Assumed input parameters. • Limits were imposed on the maximum and minimum values of parameters.
Han and Bian (2018)	SVM, ANN/ PSO	Estimating the oil recovery factor of a low permeability reservoir by using the SVM-PSO model.	• Models were for case-specific applications.
Panja et al. (2018)	ANN, LSSVM	Determining the production from shales using ML methods.	• Homogeneity in reservoir properties. • A limited number of iterations due to time constraints.

4.3. Waterflooding

Waterflooding is a common secondary recovery method because of its low cost of implementation. It involves injecting water into the reservoir to increase the production of oil. It is important to carefully design a waterflooding project to ensure that the oil recovery is achieved economically and optimally. Thus, designing a waterflooding project can be formulated as an optimization problem. In this aspect, one of the common practices of optimizing the waterflooding design is to adjust the well control rate or BHP over some time to achieve the targeted oil production that maximizes the objective function, e.g., NPV. Employing different types of algorithms to optimize waterflooding has been extensively researched in reservoir engineering. Optimization of waterflooding can induce high computational footprints especially when the investigated reservoir models are geologically complex. This is where the ML techniques have flourished as they could alleviate this computational challenge as discussed in several pieces of literature.

Mohaghegh (2011) showed that surrogate reservoir model (SRM) or smart proxy model (SPM), which represents a Neuro-Fuzzy system developed by using the database of an oil field, could be used to investigate which wells should undergo the rate constraint relaxation to ensure low water cut from waterflooding. Mohaghegh et al. (2012c) also applied SRM to a waterflooded onshore green field in Saudi Arabia to perform uncertainty quantification. Mohaghegh et al. (2012b) further extended the methodology to build well-based SRM and implemented it in two waterflooded offshore fields in Saudi Arabia for uncertainty analysis. Alenezi and Mohaghegh (2017) also successfully developed an SPM for the numerical simulation model of the waterflooded SACROC unit that accurately predicted the pressure and oil saturation values at the grid block level. To estimate the production under waterflooding, Negash and Yaw (2020) used Bayesian regularization algorithm as the training algorithm to develop an artificial neural network (ANN)-based proxy of a reservoir in Malay basin. Moreover, Zhong et al. (2020) used a more advanced ML method, conditional deep convolutional generative neural network (cDC-GAN), to build a proxy of a 2D oil-water system reservoir to forecast the field production rates under waterflooding. They also used this proxy to conduct optimization and uncertainty quantification.

Artun (2017) did a comparative study between ANN model and Capacitance Resistance Model (CRM) for the determination of interwell connectivity in waterflooded reservoirs. He stated that ANN has better flexibility in terms of modeling and data requirements since CRM is a reduced-physics model. Kalam et al. (2020) employed three approaches including non-linear regression (NLR), ANN, and adaptive neuro-fuzzy to forecast the performance of waterflooding of a stratified reservoir. They concluded that ANN yielded the best prediction. Deng and Pan (2020) also demonstrated the development of a proxy that consisted of Echo State Network (ESN) coupled with an empirical relationship of water fractional flow. This model was then used for production optimization in a closed-loop manner. SVR was also effectively employed to predict the production of a reservoir under different geostatistical realizations (da Silva et al., 2020). In another work, Bai and Tahmasebi (2020) built four different models using ANN, RNN, deep gated recurrent unit (GRU), and LSTM to predict the water coning, which has been an important issue to be handled in waterflooding. Jia and Deng (2018) used the streamline clustering AI method to identify the flowing area of waterflood in an oil field. In this work, having a reasonable number of clusters was important to have accurate clustering results. To achieve this, density peak clustering was used.

Production optimization under waterflooding of a reservoir has been frequently done with different algorithms. Guo and Reynolds (2018) developed a proxy model of a channelized reservoir by considering different geological scenarios and performed the optimization by using the stochastic simplex approximate gradient (StoSAG). Hourfar et al. (2019) employed reinforcement learning (RL) method to optimize production through waterflooding. About this, Ma et al. (2019) used deeper RL algorithms to conduct a similar optimization under geological uncertainties. They considered deep Q-network (DQN), double DQN, dueling DDQN, and deep deterministic policy gradient (DDPG). They inferred that in terms of maximization of NPV, DQN was able to perform better than the rest and as well as PSO. Furthermore, other works highlighted the useful application of metaheuristic algorithms in optimizing waterflooding. Chen et al. (2020) introduced a new methodology that was global and local surrogate-model-assisted differential evolution (GLSADE) to optimize waterflooding production. GLSADE was shown to be able to attain higher NPV than the conventional evolutionary algorithm based on three different models, such as two 100-dimensional benchmark functions, a three-channel model, and Egg model. Jia et al. (2020) suggested a data-driven optimization that included ML clustering technique and PSO for waterflooding in a complex reservoir in eastern China. ML clustering algorithm was used to identify the efficiency of waterflood performance at different layers. Then, PSO was used to conduct the optimization of the water injection plan. Peruse Table 3 for the summary of the literature on Waterflooding.

Table 3 Summary of Literature in the Domain of Waterflooding.

Literature	Methods	Remarks	Assumptions / Limitations
Mohaghegh (2011)	ANN/ GA/ Fuzzy Logic	Introducing AI-based modeling by using a case study of waterflooding.	• Models developed were case-specific.
Mohaghegh et al. (2012c)		Using AI technique to develop a Surrogate Reservoir Model (SPM) for an Onshore Green Field under waterflooding in Saudi Arabia.	
Mohaghegh et al. (2012b)		Extending the methodology to well-based SRM to two offshore fields in Saudi Arabia.	
Alenezi and Mohaghegh (2017)		Building a smart proxy model for the waterflooded SACROC unit.	
Negash and Yaw (2020)	ANN	Production prediction of the waterflooding process by using ANN.	• Existence of noise in the data collected. • Models built were case-specific.
Zhong et al. (2020)	Conditional deep convolutional generative neural network (cDC-GAN), adversarial neural network	Forecasting the field production rates of three waterflooding cases by using the neural network models.	• Limitations caused by material balance and difficulty of splitting production among producers increased the uncertainty of final results.
Artun (2017)	ANN	Implementing ANN and reduced physics	• The synthetic reservoir was set at a maximum

		model to characterize the inter-well connectivity in a waterflooded reservoir.	BHP of 5000 psia.
Kalam et al. (2020)	ANN/ Adaptive neuro-fuzzy	Estimating the oil recovery of waterflood by using AI methods in four cases: two real field cases, analytical and semi-analytical models.	• Communication between layers was assumed to be valid for the first category but not for the second. • Immiscible and piston-like displacement without gravity effects. • In this methodology, the produced water contained water coning from the injector.
Deng and Pan (2020)	Echo State Network	Embedding ML technique in Closed-Loop Reservoir Management (CLRM) for a waterflooded mature field.	• All producers were under BHP control whereas all injectors were under rate control. • The reservoir model was assumed to undergo 5 years of production before the start of the workflow. • Assumption of data acquisition frequency.
da Silva et al. (2020)	SVR	Predicting production from reservoir considering geostatistical realizations.	• The use of the dimensionality reduction method might be needed in the proposed work for much more complex cases.
Bai and Tahmasebi (2020)	LSTM	Forecasting the water breakthrough by using LSTM.	• A large variance of the training dataset.
Jia and Deng (2018)	Clustering technique	Employing streamline clustering technique to identify waterflooding flowing area in oil reservoirs.	• The flow of reservoir fluids was assumed to be along the streamline at a particular timestep.
Guo and Reynolds (2018)	SVR	Performing waterflooding optimization by using SVR-based proxy models.	• Limited total number of simulation runs for training. • The constraint of well control by simple bounds.
Hourfar et al. (2019)	RL	Applying RL to conduct waterflooding optimization.	• Voidage replacement assumption. • Operational

			constraints, like minimum and maximum injection rate, and an upper limit of the cumulative injection at each time step. • Limitation of RL: delayed reward assignment, a trade-off between exploration-exploitation, and curse of dimensionality.
Ma et al. (2019)	RL		• Assumption of production period of 1080 days. • The maximum production rate was 1500 STB/day and the minimum BHP of the producer was 1000 psi.
Chen et al. (2020)	RBF Network / DE	Conducting Global and Local surrogate modeling to optimize waterflooding with DE.	• A limited number of training data points.
Jia et al. (2020)	Machine learning algorithm/ PSO	Illustrating the combined use of ML and PSO to perform data-driven optimization of water injection plans.	• A limited number of injection plans were used.

4.4. Water-Alternating-Gas (WAG)

WAG injection is one of the most prevalent EOR techniques. It involves injecting water and gas alternately (in a cyclic manner) over a period to increase sweep efficiency to contribute to higher oil recovery. The injected gas can be CO_2 or a mixture of CO_2 and hydrocarbon gas. Optimization of WAG parameters has been widely researched because it is essential to ensure a high economic return. As stated by Mohagheghian et al. (2018), the WAG parameters generally include water and gas injection rates, BHP of producers, cycle time, cycle ratio, composition of the injected gas, total time of WAG, etc. In this context, they illustrated the successful use of metaheuristics algorithms like GA and PSO to tune the WAG parameters in Norne field to maximize the NPV and incremental recovery factor (IRF). In addition, other literature recommended the implementation of ML methods to provide fast analysis of WAG injection.

Regarding the employment of ML and metaheuristic algorithms in optimizing the WAG process, Nait Amar et al. (2018a) illustrated the development of dynamic proxy using time-dependent multi-ANN to predict the total field oil production. Then, this dynamic proxy was coupled with GA and ACO to determine the optimal WAG parameters. In addition to this, Nait Amar et al. (2020c) successfully applied SVR to build the dynamic proxy of a field in Algeria and coupled it with GA to optimize the water-alternating CO_2 gas parameters. More interestingly, the hyperparameters of SVR were optimally adjusted by GA before being used (Nait Amar et al., 2020c). Nait Amar et al. (2021) implemented two different proxies of Gullfaks field, namely Multilayer Perceptron (MLP) and Radial Basis Function Neural Network (RBFNN). Thereafter, GA and ACO were used along with these proxies to optimize the WAG process. Nwachukwu et al. (2018b) employed XGBoost to establish a proxy of a reservoir model under different geological realizations. This proxy was coupled with MADS to not only optimize the well locations but also find the optimal WAG parameters.

Belazreg et al. (2020) applied a random forest algorithm to build a model based on a database from 28 WAG pilot projects worldwide to forecast the IRF during the WAG process. Belazreg et al. (2019) also efficiently attempted the use of GMDH to develop the IRF predictive model, which was a function of horizontal and vertical permeabilities, fluid properties, mobility of fluids, WAG injection scenario, residual oil saturation to gas, trapped gas saturation, injected gas volume, and reservoir pressure. Moreover, in the work of Belazreg and Mahmood (2020), GMDH was employed to predict WAG IRF based on the data from 33 WAG projects from 28 fields in the world. Furthermore, the methodology of top-down modeling (TDM) was used by Yousef et al. (2020) to build a model to estimate the reservoir performance of a mature oil field in Middle East under WAG injection. This model also provided a rapid medium for the optimization of WAG parameters. Jaber et al. (2019b) implemented Central Composite Design (CCD) to establish a proxy of a reservoir in Subba oilfield to approximate the incremental oil recovery during the miscible CO_2-WAG process.

Nait Amar and Zeraibi (2019) established three different MLPs trained by LMA, BR, and SCG. After that, these MLPs were coupled with Non-Dominated Sorting Genetic Algorithm version II (NSGA-II) to conduct multi-objective optimization of the CO_2-WAG process. Enab and Ertekin (2020) also demonstrated how ANN could be built and used for the screening and optimization of the CO_2-WAG process and the structures of fish-bone well in low permeability reservoirs. The case study presented was a reservoir from Sirri A field. Read Table 4 for the summary of the literature on WAG.

Table 4 Summary of Literature in the Domain of WAG.

Literature	Methods	Remarks	Assumptions / Limitations
Mohagheghian et al. (2018)	GA, PSO	Optimizing WAG in Norne field with evolutionary algorithms.	• Economic constraints comprise a lower limit on oil production (10 Sm^3/day) and upper limits on water cut (0.95) and GOR (500 vol/vol). • Variables, apart from cycle ratio, cycle time, and total WAG, were assumed to be continuous.
Nait Amar et al. (2018a)	ANN/ GA, ACO,	Optimizing WAG in a synthetic field with ANN and nature-inspired algorithms.	• Imposing different constraints to the design parameters. • The database was generated based on multiple runs of the simulation.
Nait Amar et al. (2020c)	SVR/ GA	Optimizing CO_2-WAG in a synthetic field with ANN and nature-inspired algorithms.	
Nait Amar et al. (2021)	ANN/ GA, ACO	Optimizing WAG in Gullfaks field with ANN and nature-inspired algorithms.	
Nwachukwu et al. (2018b)	XGBoost/ MADS Algorithm	An extended work of Nwachukwu et al. (2018a) in which ML models were built to offer reservoir responses corresponding to well locations and control during WAG under geological uncertainty. MADS was then used for joint optimization.	• Augmentation of predictor variables due to the sophistication of response surface. • Case-sensitive application. • The proposed methodology was implemented on a synthetic case.
Belazreg et al. (2020)	Random Forest	Predictive Modeling of Incremental Recovery Factor of CO_2-WAG.	• Modeling was done based on limited/ missing data.
Belazreg et al. (2019)	GDMH. ANN	Predictive Modeling of Recovery Factor of WAG.	• WAG was assumed to begin after 10 years of waterflooding.
Belazreg and Mahmood (2020)		Predictive Modeling of WAG Incremental Recovery Factor of WAG through pilot projects.	• Modeling was done based on limited data. • The recovery factor of the pilot tests ranged from 5 to 10%.
Yousef et al. (2020)	ANN	Implementing ANN for top-down modeling in the prediction of reservoir	• A limited number of pressure tests are available. • Reservoir

		performance under WAG.	characteristics were slightly modified and assumed to be reasonably accurate for TDM. • History data (initial injection rate) was assumed to be the benchmark to assess the efficiency of injection.
Jaber et al. (2019b)	CCD	Employing a data-driven proxy to evaluate the incremental oil recovery of the CO_2-WAG process.	• 7 independent variables were assumed in the study. • The database was generated based on multiple runs of the simulation.
Nait Amar and Zeraibi (2019)	ANN/ NSGA-II	Multiobjective optimization of WAG-CO_2 in a synthetic field.	• The daily oil production rate was limited to 8500 Sm^3/day. • Total Field Oil Recovery and Total Field Water Production were assumed as objective functions. • The database was generated based on multiple runs of the simulation.
Enab and Ertekin (2020)	ANN	Applying ANN to screen and optimize CO_2-WAG and the structures of fish-bone wells in reservoirs with low permeability.	• Limitations were imposed by defining the range of each variable. • Limitations on drilling and completions were not considered.

4.5. Miscible Gas Injection

Miscible gas flooding has been one of the well-known EOR methods applied in the petroleum industry. Examples of gases usually applied in miscible gas flooding include carbon dioxide (CO_2), nitrogen (N_2), natural gas, etc. CO_2 has been preferred over other gases because implementing miscible CO_2 gas injection not only increases oil recovery but also reduces greenhouse gas emissions. Therefore, the literature survey in this section will focus mainly on miscible CO_2 gas flooding. In miscible gas injection, minimum miscibility pressure (MMP) is one of the most significant parameters that can affect the efficiency of the injection process. Accurate modeling of MMP thus has been extensively researched and application of ML in this context has also been proven successful.

Tatar et al. (2013) employed the RBFNN to estimate the MMP of pure and impure CO_2-reservoir oil. 147 data sets from different pieces of literature were used to generate the database for the modeling. Apart from RBFNN, other approaches like GA-based Backpropagation Algorithm Neural Network (GA-BPNN) were also efficiently applied by Chen et al. (2014) to develop the predictive model of MMP in the CO_2-EOR process. GA-BPNN outperformed other existing correlations as discussed in Chen et al. (2014). In addition to BPNN, Bian et al. (2016) illustrated that GA could be coupled with SVR to develop a model that could determine CO_2-oil MMP in both pure and impure streams of CO_2. GA-SVR was demonstrated to yield more accurate results of MMP than other correlations. Karkevandi-Talkhooncheh et al. (2018) used the hybrid models of RBFNN and five different metaheuristic algorithms: GA, PSO, DE, Imperialist Competitive Algorithm (ICA), and ACO. These models were able to forecast the MMP under pure and impure CO_2 injection conditions. In their study (Karkevandi-Talkhooncheh et al., 2018), ICA-RBFNN outperformed other hybrid models.

Furthermore, Nait Amar et al. (2018c) established a hybrid model of ANN and DE to forecast MMP for a pure CO_2-oil system. The initial best weight and bias parameters of ANN were optimized by employing DE. Then, this DE-optimized ANN undergoes backpropagation training again to be used as a predictive model. Nait Amar and Zeraibi (2018) also successfully tuned the hyperparameters of SVR by using ABC and applied it as a model to predict MMP in CO_2 flooding. SVR-ABC yielded a more accurate result than the SVR optimized via trial and error and other correlations. Dargahi-Zarandi et al. (2020) utilized more ML methods to develop three intelligent models to do the same prediction. These methods include Group Method of Data Handling (GMDH), MLP, and Adaptive Boosting SVR (AdaBoost SVR). Sinha et al. (2020) built four models, linear SVM, K-Nearest Neighbor regression (KNN), Random Forest Regression (RF), and ANN, to determine the MMP. They deduced that RF worked best compared with the other models. Thereafter, they substantially enhanced the RF model to become an ensemble model (hybridization of available correlation and RF) which they termed the super-learner method.

Dong et al. (2019) integrated the use of L2 regularization (which acts as a penalty term to prevent overfitting during the training phase) and dropout as a step in improving the ANN-based model that was employed to forecast MMP. This improvement could prevent the overfitting issue and further strengthen the predictive capability of the model. Other than estimating MMP, the Fuzzy Logic method was shown by Karacan (2020) capable of determining the recovery factor of miscible CO_2 gas flooding. This fuzzy-based model (with the

Mamdani-type inference system) was developed by using the data from 24 major USA field projects. You et al. (2019b) also implemented a hybrid method that considered the coupling of ANN with PSO to perform multi-objective optimization of CO_2-EOR. The objective functions included CO_2 storage, oil recovery factor, and NPV. The literature on Miscible Gas Injection is summarized in Table 5.

Table 5 Summary of Literature in the Domain of Miscible Gas Injection.

Literature	Methods	Remarks	Assumptions / Limitations
Tatar et al. (2013)	RBFNN	Modeling of the CO_2-reservoir oil minimum miscibility pressure.	• Models were developed based on available experimental data.
Chen et al. (2014)	Backpropagation Neural Network / GA		
Bian et al. (2016)	SVR / GA	Modeling of CO_2-oil minimum miscibility pressure with pure and impure CO_2.	• Built based on available experimental data. • Separate models for pure and impure CO_2.
Karkevandi-Talkhooncheh et al. (2018)	Radial Basis Function Networks / GA, PSO, ICA, ACO, DE		• Models were extended on a basis of limited data points.
Nait Amar et al. (2018c)	ANN / DE	Developing the predictive model of minimum miscibility pressure in a pure CO_2-oil system.	• Models were built based on data from a few experiments. • Choice of input parameters was assumed.
Nait Amar and Zeraibi (2018)	SVR / ABC	Building the predictive model of minimum miscibility pressure in the CO_2-EOR process.	
Dargahi-Zarandi et al. (2020)	Adaptive Boosting SVR, GDMH, MLP	Predictive Modeling of MMP of pure and impure CO_2-crude oil systems.	• Predicting the limited range of MMP between 1000 psia and 4900 psia. • Dataset limitation.
Sinha et al. (2020)	Linear SVM/ K-Nearest Neighbor Regression/ Random Forest regression/ ANN	Predictive Modeling of MMP of CO_2-crude oil systems.	• Data set limitation. • Further applicability of models was limited.
Dong et al. (2019)	ANN		• A limited number of field cases. • Input variables were assumed based on the availability of data.
Karacan (2020)	Fuzzy Logic	Forecasting of recovery factor of miscible CO_2-EOR.	• The model was constructed by only using data from 24 U.S. field projects.

You et al. (2019b)	ANN/ MO-PSO	Applying ANN for multi-objective optimization of CO_2-EOR.	• Only 4 input parameters: water cycle, gas cycle, BHP of producer, and water injection rate, were considered.

4.6. Other EOR Techniques

EOR methods can be fathomed as tertiary recovery techniques used to retrieve the remaining oil from hydrocarbon reservoirs. These techniques will be initiated after the exhaustion of both primary and secondary recovery methods. Examples include surfactant flooding, polymer flooding, any other chemical flooding, nitrogen gas injection, in-situ combustion, Steam-Assisted Gravity Drainage (SAGD), cyclic steam injection, fire-flooding, microbial flooding, and so forth. The cost of implementation of these methods is relatively higher than primary and secondary recovery methods. Therefore, careful design and optimization of the tertiary recovery methods are important to elude any unnecessary waste of expenditure and ensure the profitability of the project. Several works have illustrated the application of ML methods in the context of the employment of different tertiary recovery techniques.

Rezaian et al. (2010) applied experimental methods to examine the effect of Poly Vinyl Acetate (PVA) on the rheology of crude oil and water. This was because they wanted to study the effectiveness of PVA to be used in polymer flooding. Thereafter, they demonstrated the successful implementation of ANN in developing a predictive model based on the experimental data. Zerafat et al. (2011) illustrated the use of Bayesian network analysis to screen for an efficient EOR method. They applied different data sets from seven different EOR methods, like miscible N_2 injection, miscible hydrocarbon injection, miscible and immiscible CO_2 injection, polymer flooding, in-situ combustion, and steam injection. Siena et al. (2016) further built a novel EOR screening tool by using the Bayesian approach. The approach they implemented included Bayesian Hierarchical Clustering (BHC) algorithm and PCA, which could be understood as a two-step algorithm. PCA was used to reduce the dimensionality of data and provide accurate distance metrics regarding the similarity among the projects. The database they used generally comprised thermal EOR, chemical EOR, and gas/WAG injection that were derived from different worldwide projects and literature.

Parada and Ertekin (2012) applied ANN modeling to establish a new screening tool for four different recovery methods including waterflooding, miscible N_2 injection, miscible CO_2 injection, and steam injection. Khazali et al. (2019) presented the use of a fuzzy decision tree in the assessment of EOR screening. They stated that the fuzzy decision tree could perform the simultaneous ranking and classification of different EOR techniques. Hence, an expert system could be designed to generate the EOR rules. In their work, the decision tree was applied to the dataset of 548 observations related to ten different EOR methods. Sun and Ertekin (2020) showed that ANN-based proxies could be established to do the screening of polymer flooding. Then, they coupled the proxies with PSO to optimize the polymer flooding process to maximize the NPV. In the domain of optimization, Ma and Leung (2020) designed a hybrid workflow that integrated multi-objective optimization (MOO) and proxy modeling in the case of injection of warm solvent into heterogeneous heavy oil reservoirs. In their work (Ma and Leung, 2020), NSGA-II was used to perform the MOO.

Regarding recovery performance forecasting, Ehsan et al. (2014) applied PCA to decrease the dimensionality of the input data before modeling the ANN. The ANN was used to estimate the production induced by the SAGD process in heterogeneous reservoirs. Ersahin and Ertekin (2020) also conducted the development of ANN of cyclic steam injection (CSI) in naturally fractured reservoirs. The ANN models developed included a forward model and two inverse models. The forward model was used to estimate the cumulative oil production and changes in viscosity near the wellbore. About the inverse-looking models, the first one was used to find out the ideal design of injection variables whereas the second one was used for the characterization of some reservoir properties. Abdullah et al. (2019) developed five ANN models to be implemented in chemical EOR in a sandstone reservoir. These models were applied to estimate reservoir performance, forecast reservoir properties, determine the design parameters for known performance and properties, and find out the design parameters for a targeted cumulative oil production and project period. Refer to Table 6 for the summary of the literature on other EOR techniques.

Table 6 Summary of Literature in the Domain of Other EOR Techniques.

Literature	**Methods**	**Remarks**	**Assumptions / Limitations**
Rezaian et al. (2010)	ANN	ML models were built to predict the effect of Poly Vinyl acetate on the rheology of water and crude oil in EOR.	• Data was only from one experiment and this might limit the applicability of the models developed. • The experiment was done under predefined conditions.
Zerafat et al. (2011)	Bayesian Network	The model was created as a tool for EOR screening based on	• The study was done without considering economic limitations.

		data from 10 Iranian southwest reservoirs.	• Models were case-specific.
Siena et al. (2016)	Bayesian Clustering/ PCA	A novel EOR screening tool was established.	• Evaluation of probability based upon the fundamental assumption of Bayesian clustering. • Identification of analogs is vital to the successful implementation of this methodology.
Parada and Ertekin (2012)	ANN	An ANN-based EOR screening tool was built.	• Ability to predict reservoir response to different conditions within certain limits. • Four different compositions of hydrocarbon were considered.
Khazali et al. (2019)	Fuzzy Decision Tree	EOR screening evaluation by using a fuzzy decision tree.	• The proposed method works best with sufficient data. • Economic issues were not concerned.
Sun and Ertekin (2020)	ANN	The ANN-based model was created to screen and optimize polymer flooding.	• Salinities of injected and in-situ water were assumed the same. • Gravitational forces and capillary pressure were assumed to be negligible. • Existence of upper and lower limits of the search space of design parameters.
Ma and Leung (2020)	ANN/ NGSA-II	Hybridization of ANN and NSGA-II for multi-objective optimization of warm solvent injection in heterogeneous heavy oil reservoirs.	• Assumption of uniform properties within each facies. • Only sand was assumed to exist at the well grid cell. • Only bottom-hole pressures were chosen as design parameters. • Excessive startup time and slow extraction rate limited the application.
Ehsan et al. (2014)	ANN/ PCA	An integrated approach of ANN and PCA for the prediction of SAGD performance in heterogeneous reservoirs.	• The study was limited to the database that was created from the combinations of the attributes of heterogeneous reservoir as input. • Separate ANNs were required for better results.

Ersahin and Ertekin (2020)	ANN	Using ANN to model the Cyclic Steam Injection Process in Naturally Fractured Reservoirs.	• Oil behaves as Newtonian fluid. • A trial-and-error approach was needed to train the ANN and determine its optimum design.
Abdullah et al. (2019)	ANN	Applying ANN to design and model the implementation of chemical EOR.	• The surfactant was in the aqueous phase. • Data available was assumed to be reservoir characteristics, project duration aimed, and cumulative oil volume.

4.7. Carbon Capture and Storage (CCS)

The increasing amount of carbon dioxide (CO_2) gas in the atmosphere is one of the main factors contributing to climate change today. Nevertheless, CO_2 emission is an inevitable consequence of different types of industrial and commercial activities required to fulfill our daily practical needs. Therefore, awareness has arisen among researchers to look for an efficient strategy to reduce CO_2 emissions. One of the proposed strategies to assure that emission of CO_2 will remain at a low level is Carbon Capture and Storage (also known as Carbon Capture and Sequestration) (CCS). Fundamentally, CCS is performed by injecting the captured CO_2 into geological formations and ensuring it is safely trapped underground. Much research has been done on the domain of CCS and one of the most cutting-edge topics is the coupling of ML techniques with CCS. Several pieces of literature also discussed the application of metaheuristic algorithms along with the ML methods in CCS.

Sipöcz et al. (2011) developed two different ANN models to predict the CO_2 capturing processes. The difference between the models was the training algorithm used where one was trained using scaled conjugate gradient (SCG) algorithm whereas the other training algorithm employed was Levenberg Marquardt algorithm (LMA). They deduced that these models could provide results not only much faster than process simulator CO2SIM but also within an acceptable level of accuracy. Miscibility of CO_2 in formation fluids is another important aspect of CCS. Mesbah et al. (2018) illustrated the implementation of a multilayer perceptron neural network (MLP-NN) by employing 1386 experimental data points to forecast the miscibility of CO_2 and supercritical CO_2 in ionic liquid. During the development of the model, they performed outlier diagnostics to ensure the quality of data used. Furthermore, Sinha et al. (2020) used ML methods, like random forest and multilayer feedforward neural network (MFNN), to build models for leakage detection in a carbon sequestration project in Cranfield reservoir, Mississippi, USA. The models were made based on time series signals from the pressure pulse test. Thanh et al. (2020) also successfully showed the use of ANN to estimate the performance of CO_2-EOR and storage in a residual oil zone located in Permian basin.

Metaheuristic algorithms were also proven to be useful to be coupled with ML techniques in CCS. You et al. (2019a) provided a framework to conduct co-optimization on CO_2 storage, the performance of CO_2-EOR, and the NPV of the project. In the framework, RBFNN and multilayer neural network modeling were implemented to build the proxies of the reservoir model. Then, PSO was used to do the co-optimization. After that, You et al. (2020c) also developed ANN to establish a proxy of the sandstone reservoir in Pennsylvanian Upper Morrow to estimate the time series of cumulative oil production and CO_2 storage. PSO was again applied to co-optimize CO_2 storage, the performance of CO_2-EOR, and the NPV of the project. In addition to proxy modeling, other interesting literature have discussed the use of ML to predict important parameters relevant to CCS. The solubility of CO_2 in formation fluid is an essential parameter to be considered in CCS. In this context, Nait Amar et al. (2019) applied MLP and RBFNN to make predictive models of CO_2 solubility in brine. More intriguingly, LMA was employed to train MLP whereas GA, ABC, and PSO were used to train RBFNN. In their study, RBFNN-ABC outperformed the other models. Hemmati-Sarapardeh et al. (2020a) also used four ML techniques, including RBFNN, MLP, Least-Squares Support Vector Machine (LSSVM), and Gene Expression Programming (GEP), to model the solubility of CO_2 in water at high temperature and pressure. During the training phase, four backpropagation algorithms were used in the modeling of MLP whereas four nature-inspired algorithms were used in the modeling of RBFNN and LSSVM. These nature-inspired algorithms included PSO, GA, FA, and DE.

In addition, Nait Amar and Jahanbani Ghahfarokhi (2020) presented how white-box ML methods could be used to estimate CO_2 diffusivity in brine. These white-box ML techniques were GMDH and GEP. These models could be applied to predict the diffusivity coefficient of CO_2 in brine as functions of temperature, pressure, and viscosity of the solvent. Also, Nait Amar et al. (2020a) utilized MLP, GMDH, and GEP to build predictive models of CO_2 viscosity at high temperature and pressure. Four backpropagation algorithms, LMA, SCG, Bayesian Regularization (BR), and Resilient Backpropagation (BR), were used to train the MLP. The thermal conductivity of carbon dioxide is another important parameter in CCS projects. Regarding this, Nait Amar et al. (2020b) first established some MLP-based models and RBFNN trained by PSO to forecast the thermal conductivity of carbon dioxide. After that, the two best models were coupled with two Committee Machine

Intelligent Systems (CMIS) via the weight averaging method and GMDH. Peruse Table 7 for the summary of the literature on CCS.

Table 7 Summary of Literature in the Domain of CCS.

Literature	Methods	Remarks	Assumptions / Limitations
Sipöcz et al. (2011)	ANN	ML was employed for the modeling and prediction of the CO_2 capture process plant.	• Limited to 5000 epochs due to low computational space. • Each input parameter was assumed and underwent sensitivity analysis to assess its dependence on the output.
Mesbah et al. (2018)	MLP	ML was used to develop predictive models of miscibility of CO_2 and supercritical CO_2 in ionic liquid.	• Input parameters used for modeling were assumed. • The methodology is yet subject to the verification of other databases.
Sinha et al. (2020)	Multilayer FNN. Random Forest, Linear models	ML models were established for leakage detection in a carbon Sequestration project.	• Simplistic ML techniques showed limited sufficiency in capturing the details. • The window of 1000 samples was not decided through a comprehensive analysis.
Thanh et al. (2020)	ANN/ PSO	ANN was applied to forecast the performance of CO_2 EOR and storage in a residual oil zone.	• The model developed is case-specific. • The selection of the range of uncertainty parameters requires more attention.
You et al. (2019a)	RBFNN, Multilayer Neural Networks / PSO	An optimization framework, considering ML and PSO, was proposed to co-optimize CO_2 EOR and storage in a sandstone reservoir.	• Production pressure is limited to 4000 psia whereas injection pressure is 5000 psia. • Three different development strategies were assumed.
You et al. (2020c)	ANN / PSO	A part of the extended work of You et al. (2019a). An ML-assisted computational workflow was introduced to optimize a CO_2-WAG injection plan that considers CO_2 sequestration and hydrocarbon recovery.	• Production pressure is limited to 4000 psia whereas injection pressure is 5000 psia. • Operational cost is primarily influenced by the amount of CO_2.

Nait Amar et al. (2019)	MLP, RBFNN / GA. PSO, ABC	Different ML models were built to determine the solubility of CO_2 in brine, which is important to the application of CCS.	• Limited to the database used for modeling (robustness still needs to be verified). • Input data parameters were assumed for developing the models
Hemmati-Sarapardeh et al. (2020a)	LSSVM, GEP / PSO, GA, DE, FA	Numerous ML methods were implemented to estimate the solubility of CO_2 in water at high pressure and temperature.	
Nait Amar and Jahanbani Ghahfarokhi (2020)	GMDH, GEP, Decision Trees, Random Forests.	Models that could forecast CO_2 diffusivity in brine were established with the aid of ML.	
Nait Amar et al. (2020a)	MLP, GEP, GDMH	Numerous ML methods were employed to predict the viscosity of CO_2 at high pressure and temperature.	
Nait Amar et al. (2020b)	MLP, RBFNN, CMIS, CMIS-GMDH	Models that could forecast CO_2 thermal conductivity were established with the aid of ML.	

4.8. History Matching

History Matching (HM) can be understood as a task that involves tuning or adjustment of any parameter that is used in reservoir modeling to enable a reservoir model to yield results that match the observed real-field data. It can be understood that HM can be very laborious and time-consuming. To mitigate this computational challenge, several works propose the application of ML techniques in establishing the proxies of the numerical reservoir models to be employed in HM. Besides that, HM is considered an optimization problem as it involves the minimization of the error between the predicted data and observed data. In this aspect, metaheuristic algorithms have widely contributed to the successful and efficient deployment of HM. More intriguingly, some literature highlighted the coupling of proxies with metaheuristic algorithms in performing HM. Thus, ML and metaheuristic algorithms show great potential to be further improved in the future implementations of HM.

Sampaio et al. (2009) presented the fundamental use of FNN as the nonlinear proxy model of a numerical and synthetic heterogeneous model. Then, they applied it in HM and showed very positive results. However, they opined that the complexity of the reservoir model could be increased to illustrate the robustness of ML. Shahkarami et al. (2014b) proposed the use of a surrogate reservoir model (SRM), which was represented as a Neuro-Fuzzy system, in the HM phase. They termed it AI-assisted HM (AHM) and successfully showed that it could reduce the computational time induced by the conventional approach of HM using a very heterogeneous model. Masoudi et al. (2020) employed a similar methodology to conduct HM on a very complicated and mature offshore oilfield in Malaysia. However, the SRM used was the deconvolutional neural network. Also, they applied top down modeling (TDM) that included the data from the real field in designing the SRM used for HM. Illarionov et al. (2020) studied different approaches to HM of a real-field model on an FNN-based proxy termed as Neural Differential Equations based Reduced Order Model (NDE-b-ROM). The HM methods considered a variation of reservoir model parameters, an adaptation of neural network architecture, and an adaptation of latent space of model parameters. They inferred that latent space adaptation would yield the best result.

More advanced techniques were also used in proxy modeling along with HM. Chaki et al. (2020) employed deep neural networks (DNN) and RNN to build proxy models of Brugge reservoir and conducted an exhaustive search of HM using the models. Honorio et al. (2015) also included a novel ML method to study the prior information on geology and use pluri-principal-component-analysis (pluri-PCA) to rebuild a model. Fundamentally, they implemented pluri-PCA to transform the geological models to Gaussian PCA coefficients and tuned them in HM. Rammay et al. (2020) examined different algorithms used for HM of imperfect subsurface models. These algorithms included HM without considering model error, HM with an update of total error covariance matrix through iteration, HM with PCA-based error model, HM with PCA-based error model and noise covariance matrix, HM with PCA-based error model and considering second-order errors, and HM with PCA-based error model and update of total error covariance matrix through iteration. They deduced that the last three algorithms yielded models with high fidelity. Liu and Durlofsky (2020) also illustrated the use of optimization-based PCA (O-PCA) and CNN-based PCA as geological parametrization techniques to represent the model properties of complex reservoirs. These techniques were coupled with the MADS to do HM. Also, the proxy-based Markov Chain Monte Carlo algorithm was successfully employed with the Embedded Discrete Fracture Model (EDFM) to conduct AHM on the oil well in Vaca Muerta shale (Dachanuwattana et al., 2018). An ensemble smoother neural network (ES-NN) that comprised ensemble smoother (ES) and convolutional autoencoder (CAE) was built and used to HM the channelized reservoirs by Kim et al. (2020). They stated that the ES-NN produced better performance than the ensemble smoother-multiple data assimilation (ES-MDA).

As discussed before, metaheuristic algorithm has been efficiently proven successful as an optimization algorithm in HM. Schulze-Riegert et al. (2002) applied evolutionary algorithms to conduct HM of a sophisticated synthetic reservoir model of a North Sea reservoir. Karimi et al. (2017) used GA along with the proxy model, which was the RSM of a 3D giant reservoir model, to do HM. Kriging proxy modeling and Sobol sampling sequence were applied by Shams et al. (2019) to do AHM by implementing three metaheuristic algorithms such as Firefly Optimization (FFO), Bee Colony Optimization (BCO), and Harmony Search Optimization (HSO). Shahkarami et al. (2018) illustrated the AHM by implementing the technology of pattern recognition. They established SRM of PUNQ-S3 reservoir model by applying ten realizations and coupled the SRM with DE to perform the AHM. In addition, He et al. (2016) applied a similar methodology to develop a proxy model of SACROC unit (Scurry Area Canyon Reef Operational Committee) which was the main part of the Kelly Snyder field in the Permian Basin. They also successfully coupled the proxy with DE to do AHM. Riazi et al. (2016) demonstrated the use of LSSVM to develop a proxy model of a fractured reservoir. Thereafter, they successfully implemented PSO and ICA to do AHM. Rana et al. (2018) suggested applying Gaussian Process-based Proxy Modeling and Variogram-based Sensitivity Analysis (GP-VARS) on the PUNQ-S3 reservoir to solve the HM problem. They mentioned that this methodology was four times computationally less demanding than using DE on the numerical simulation to do HM. The literature on History Matching is summarized in Table 8.

Table 8 Summary of Literature in the Domain of History Matching.

Literature	Methods	Remarks	Assumptions / Limitations
Sampaio et al. (2009)	FNN	Using FNN to perform History Matching.	• Input parameters of FNN were assumed. • The size of the training group was assumed. • The simplicity of the case study.
Shahkarami et al. (2014b)	FNN/ Fuzzy Logic	Implementing SRM in the workflow of AI-Assisted History Matching in a synthetic but heterogeneous reservoir model.	• Models were meant for case-specific applications. • A limited number of uncertain variables.
Masoudi et al. (2020)	Deconvolutional Neural Networks	Applying TDM to conduct History Matching in a highly sophisticated field in Malaysia.	• Models were meant for case-specific applications. • No guideline on determining the sequence of separate

			TDMs.
Illarionov et al. (2020)	FNN	Doing gradient-based History Matching with the help of FNN on a field model.	• Assumption of limited prior knowledge of the geological parameters. • Adaptation of the workflow concerning production rates data was not considered.
Chaki et al. (2020)	Deep Neural Network / RNN	Performing History Matching on the Brugge field model.	• Testing of the suggested methodology was required for a more complex reservoir model. • Limited input parameters were considered.
Honorio et al. (2015)	Piecewise Reconstruction from a Dictionary (PRaD)/ pluri-PCA	Developing an assisted History Matching with PRaD and pluri-PCA based on a case study of geologically complex reservoirs.	• Assumption of independence of measurement errors.
Rammay et al. (2020)	PCA-based error model	Integrating different approaches to the PCA-based error model in the History Matching workflow on a case study.	• Limiting the test setting, viz. coarsened grids and upscaled geological feature, to cases with the availability of high-fidelity model with model discrepancy from the historical data. • The limited capacity of the error model.
Liu and Durlofsky (2020)	CNN-based PCA/ MADS	Proposing the use of CNN-PCA for geological parameterization in the workflow of History Matching.	• Facies types and log-permeability values were assumed to be available in one of the case studies. • Random noise was assumed to be independent. • Assumption of full parallelization.
Dachanuwattana et al. (2018)	K-NN algorithm/ Markov Chain Monte Carlo (MCMC)	Demonstrating the use of K-NN-based and MCMC-based proxies to history match a shale oil well.	• Assumption of uniform distribution of uncertain parameters. • Production of the well was assumed to be at a BHP of 500 psi for 8000 days. • Uniform distribution of fractures.
Kim et al. (2020)	Ensemble Smoother-Neural Network (ES-NN)	Presenting the use of ES-NN in the workflow of History Matching.	• Assuming the time of measurement of observation data during History Matching.

			• Each facies was assumed to have a constant permeability value.
Schulze-Riegert et al. (2002)	Evolutionary Algorithm in a Multi-purpose Environment for Parallel Optimization (MEPO)	Illustrating the application of the evolutionary algorithm in the context of MEPO in the history matching of a complex black oil model.	• Multi-dimensional search space was assumed for the reservoir studied. • Unavailability of information on reservoir beyond geostatistical, geological, seismic, and history data. • Independence of measurement errors. • Parameters were correlated in the region identified. • Homogeneous porosity of 0.30.
Karimi et al. (2017)	Genetic Algorithm	Incorporating GA in the History Matching with the use of a proxy model.	• The proposed methodology might not be computationally favorable with reservoirs of more than 20 wells.
Shams et al. (2019)	ANN/ GA, PSO, Firefly Algorithm, Bee Colony, Harmony Search	Introducing the use of 3 nature-inspired algorithms in the History Matching along with a proxy model.	• Models were meant for case-specific applications.
Shahkarami et al. (2018)	FNN/ DE	Presenting the coupling of SRM and DE for History Matching in PUNQ-S3.	• Reservoir properties were assumed to be measured at well locations. • Models were meant for case-specific applications.
He et al. (2016)		Presenting the coupling of SRM and DE for History Matching in the SACROC unit.	• Models were meant for case-specific applications.
Riazi et al. (2016)	LSSVM/ PSO, ICA	Establishing the LSSVM-based proxy and coupling it with the algorithms for History Matching in a fractured reservoir.	• Properties of fractures were assumed to be homogeneous.
Rana et al. (2018)	Gaussian Process proxy / Variogram-based sensitivity analysis	Illustrating the efficient assisted History Matching with Gaussian Process proxy in PUNQ-S3.	• Lacking validation of the proposed workflow in a more complex reservoir.

5. PROS, CONS, AND OTHER DISCUSSIONS

5.1. Pros

As briefly mentioned, one of the main advantages of applying ML-based approaches in the context of reservoir simulation, is the reduction of computational footprints. Even with the current improvements in computational power, numerical simulation of a very sophisticated reservoir model may take a few months in field development studies. Therefore, it is essential to find alternatives that can speed up the calculation. This is where the intelligent proxy can contribute. If an intelligent proxy model with high fidelity is successfully established, any decision problem related to reservoir management can be handled much more quickly. Thus, further inconvenience can be avoided especially when any relevant reservoir management plan needs to be updated at a high frequency.

In addition, the mechanism of the ML-based methods is very comprehensible as it generally does not involve complicated mathematical equations. Hence, when it comes to application, we believe that it will not pose any additional challenges. Albeit there are some contretemps mentioning that ML-based methods are "black-box", we do not completely abide by this opinion as we think the formulations of ML-based approaches are not as opaque as claimed. Fundamentally, these methods are explainable through mathematics. For instance, the mechanism of ANN is established by treating the nodes as neurons in the human brain. Thereafter, the weights and biases which connect the nodes in different layers are continuously adjusted using any algorithm to enable the ANN to achieve learning. From this, if we can perceive how the ML-based methods work mathematically, the implementation should be convenient. Another benefit of implementing the ML-based models, particularly in the case of TDM, pertains to the exclusion of assumptions and simplifications of physics. This is different from applying the physics-based models that might still require a few assumptions to forecast the production from a reservoir which can be problematic in dealing with real field data. In other words, the complex physics of the system might not be captured well with assumptions. In this context, data acts as a guide to the solution.

Based on the previously discussed literature, the petroleum industry is gradually gaining maturity in this domain of technology. ML-based methods offer high robustness in terms of application. Robustness here indicates that these methods can generally solve any kind of engineering problem if the problem is well-formulated, and the data are properly prepared. Aside from reservoir engineering, the use of ML-based methods in drilling engineering (Barbosa et al., 2019; Mahmoud et al., 2021; Tunkiel et al., 2020), production engineering (Huang and Chen, 2021; Wei et al., 2021; Zhong et al., 2020), petrophysics (Ali et al., 2021; Blanes de Oliveira and de

Carvalho Carneiro, 2021; Osarogiagbon et al., 2020), etc. has been successful. Thus, they have been termed panacea for most problems. We would like to emphasize that the use of ML-based methods ought to be upheld but should not be treated as the only solution. In this case, we refer to the hand-shaking protocol proposed by Ertekin and Sun (2019).

5.2. Cons

There are also some limitations associated with the use of ML-based methods. One of them includes long training time caused by a large database. One needs to consider a trade-off between the size of the database and training time when he or she plans to build intelligent proxies. We reckon that creativity is required in the phase of problem formulation to avoid long training time in the later stages. The benefit of intelligent proxy modeling is better demonstrated in the case of very complicated and heterogeneous reservoir models in which the simulation time would exceed that of neural network training by certain orders of magnitude (Mohaghegh, 2017a; Shahab Dean Mohaghegh, 2011). This implies that having an intelligent proxy for a simplistic case does not showcase its real potential. However, having an intelligent proxy to capture a sophisticated physical relationship is noteworthy. The overfitting issue is another problem that needs to be dealt with when ML-based methods are applied. If the intelligent proxy is not well-trained, the data partitioning and training will have to be repeated. Mitigating overfitting can be laborious depending upon the complexity of the database. In addition, building intelligent proxies requires a very clear objective. Thus, it is not a one-size-fits-all model. This limitation may hinder some reservoir engineers from tending to attempt intelligent modeling. In terms of modeling with real field data, only the database from a brown field is deemed reliable in developing a useful DDM. This is because the amount of data should be sufficiently big to reflect the physics of fluid flow throughout a long period of production. In this case, another limitation also arises where there might be some missing data points during the collection of real field data for establishing a DDM. Hence, as recommended by Mohaghegh (2005), a viable solution is doing statistical averaging.

5.3. Other Discussions

Our survey also touched upon the application of metaheuristic algorithms along with intelligent models. Several studies (Nait Amar et al., 2018b, 2018c) proposed that when the metaheuristic algorithms are followed by conventional backpropagation algorithms in neural network training, the respective ANN illustrates better predictability. In addition, for intelligent proxy modeling, implementing metaheuristic algorithms is relatively more explicit and convenient than employing the derivative-based approaches because these algorithms do not require the approximation of the gradient. Therefore, applying them can be convenient if the corresponding mechanism can be mathematized accordingly. Ezugwu et al. (2020) illustrated the benefits and drawbacks of applying 12 metaheuristic algorithms: Cuckoo Search, DE, GA, PSO, Symbiotic Organism Search, FA, ACO, Bat Algorithm, Flower Pollination Algorithm, ABC Algorithm, Bee Algorithm, and Inverse Weed Optimization. In general, most of these algorithms have a better ability to converge to the global optimum whereas some of them might have low convergence rates and yield partially optimal results. Therefore, it is recommended to understand both the advantages and disadvantages of any chosen metaheuristic algorithm before employment. According to our studies, metaheuristic algorithms illustrate a very huge potential to be extensively applied in different domains of reservoir engineering.

To further generalize the application of ML-based methods in reservoir engineering, especially in the intelligent proxy modeling of NRS, we have summarized a few areas which might need more scrutiny. The first area is the sampling strategies. Our investigation reveals that a more efficient sampling method can be used to enable the development of more robust intelligent proxies. The efficiency of the sampling methods is defined as its ability to retrieve samples that can cover the solution space as extensively as possible. In this aspect, we opine that the coupling of two different sampling strategies, namely Latin Hypercube method and Sobol sequence, as initiated by Dige and Diwekar (2018), can be treated as an alternative to assess whether a better intelligent proxy can be developed. Exploring any better feature selection method to mitigate the curse of dimensionality is also thinkable. About this, Mohaghegh (2017a) has initiated the application of fuzzy pattern recognition in selecting more useful input variables in proxy modeling. However, we reckon that other approaches, viz. mutual information method based on Shannon entropy in information theory (Shannon, 1948; Thanh et al., 2022), can be considered to verify whether improvement can be achieved. We would like to emphasize that if an intelligent proxy is developed upon the results of a numerical model, this proxy can only act as a complement. This is because the source of data is the NRS. When the data come from real field measurements, it is however a research question to investigate whether the intelligent proxy can completely replace the NRS in solving reservoir management-related problems. To the best of our knowledge, there are not many studies that discuss the development of coupled ML-metaheuristic paradigm while there are numerous discussions regarding the separate use of ML and metaheuristic algorithms. We hope that this survey can provide insights to the research community to further explore the potential of coupled ML-metaheuristic paradigm in the context of reservoir engineering.

6. SUMMARY

In this work, we have surveyed the employment of ML methods and coupled ML-metaheuristic paradigm in developing proxies of numerical simulation models where only a limited number of literature studied the latter. Nevertheless, the respective literature were included along with other articles that mainly touched upon the employment of ML in the domain of reservoir simulation to highlight the robustness of ML methods. We illustrated the general framework and several suggestions, including proper identification of the objective of proxies and data normalization, that could be implemented to successfully develop an intelligent proxy model. Albeit these recommendations seemingly appear to be trivial, it happens that they could have been overlooked in proxy modeling. In addition, we demonstrated and discussed the application of ML approaches and the hybrid approach in different domains of reservoir engineering such as well placement, monitoring production parameters, miscible gas injection, waterflooding, CCS, WAG, other EOR methods, and history matching. We also briefed on the pros and cons of using ML approaches and metaheuristic algorithms. We opined that several aspects associated with intelligent proxy modeling need to be addressed to achieve further maturity in the application of this technology. In general, we can infer that the ML methods and the coupled paradigm provide useful insights into the resolution of reservoir management issues. Furthermore, ANN is portrayed as very flexible to be implemented to build intelligent proxies. Therefore, despite not being a "one-size-fits-all" solution, these methods ought to be further explored due to their huge potential. We also conclude that the potential of coupled ML-metaheuristic paradigm can still be further investigated mostly in the context of reservoir simulation. This survey paper aims at inspiring and providing insights for other researchers and engineers concerning this.

Acknowledgement

This research is a part of BRU21 – NTNU Research and Innovation Program on Digital Automation Solutions for the Oil and Gas Industry (www.ntnu.edu/bru21).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abdullah, M., Emami-Meybodi, H., Ertekin, T., 2019. Development and application of an artificial neural network tool for chemical EOR field implementations, in: Society of Petroleum Engineers - SPE Europec Featured at 81st EAGE Conference and Exhibition 2019. https://doi.org/10.2118/195492-ms

Afzali, S., Rezaei, N., Zendehboudi, S., 2018. A comprehensive review on Enhanced Oil Recovery by Water Alternating Gas (WAG) injection. Fuel. https://doi.org/10.1016/j.fuel.2018.04.015

Ahmadi, M.A., Pouladi, B., Barghi, T., 2016. Numerical modeling of CO2 injection scenarios in petroleum reservoirs: application to CO2 sequestration and EOR. J. Nat. Gas Sci. Eng. 30, 38–49.

Ahmadi, M.A., Zendehboudi, S., James, L.A., 2018. Developing a robust proxy model of CO 2 injection: Coupling Box--Behnken design and a connectionist method. Fuel 215, 904–914.

Ahmed, T., 2018. Reservoir engineering handbook. Gulf professional publishing.

Al-Alwani, M.A., Britt, L., Dunn-Norman, S., Alkinani, H.H., Al-Hameedi, A.T., Al-Attar, A., 2019. Production performance estimation from stimulation and completion parameters using machine learning approach in the Marcellus shale, in: 53rd U.S. Rock Mechanics/Geomechanics Symposium.

Alakeely, A., Horne, R.N., 2020. Simulating the Behavior of Reservoirs with Convolutional and Recurrent Neural Networks, in: SPE Reservoir Evaluation and Engineering. https://doi.org/10.2118/201193-PA

Alatrach, Y., Mata, C., Omrani, P.S., Saputelli, L., Narayanan, R., Hamdan, M., 2020. Prediction of well production event using machine learning algorithms, in: Society of Petroleum Engineers - Abu Dhabi International Petroleum Exhibition and Conference 2020, ADIP 2020. https://doi.org/10.2118/202961-ms

Alenezi, F., Mohaghegh, S., 2017. Developing a Smart Proxy for the SACROC Water-Flooding Numerical Reservoir Simulation Model, in: SPE Western Regional Meeting Proceedings. https://doi.org/10.2118/185691-ms

Ali, M., Jiang, R., Huolin, M., Pan, H., Abbas, K., Ashraf, U., Ullah, J., 2021. Machine learning - A novel approach of well logs similarity based on synchronization measures to predict shear sonic logs. J. Pet. Sci. Eng. 203, 108602. https://doi.org/10.1016/j.petrol.2021.108602

Alkhalaf, A., Isichei, O., Ansari, N., Milad, R., 2019. Utilizing machine learning for a data driven approach to flow rate prediction, in: Society of Petroleum Engineers - Abu Dhabi International Petroleum Exhibition and Conference 2019, ADIP 2019. https://doi.org/10.2118/197266-ms

Alkinani, H.H., Al-Hameedi, A.T.T., Dunn-Norman, S., Flori, R.E., Alsaba, M.T., Amer, A.S., 2019. Applications of artificial neural networks in the petroleum industry: A review, in: SPE Middle East Oil and Gas Show and Conference, MEOS, Proceedings. https://doi.org/10.2118/195072-ms

Amaechi, U.C., Ikpeka, P.M., Xianlin, M., Ugwu, J.O., 2019. Application of machine learning models in predicting initial gas production rate from tight gas reservoirs. Rud. Geol. Naft. Zb. https://doi.org/10.17794/rgn.2019.3.4

Ani, M., Oluyemi, G., Petrovski, A., Sina, R.G., 2016. Reservoir uncertainty analysis: The trends from probability to algorithms and machine learning, in: Society of Petroleum Engineers - SPE Intelligent Energy International Conference and Exhibition. https://doi.org/10.2118/181049-ms

Artun, E., 2017. Characterizing interwell connectivity in waterflooded reservoirs using data-driven and reduced-physics models: A comparative study. Neural Comput. Appl. https://doi.org/10.1007/s00521-015-2152-0

Bai, T., Tahmasebi, P., 2020. Efficient and data-driven prediction of water breakthrough in subsurface systems using deep long short-term memory machine learning. Comput. Geosci. https://doi.org/10.1007/s10596-020-10005-2

Balaji, K., Rabiei, M., Suicmez, V., Canbaz, H., Agharzeyva, Z., Tek, S., Bulut, U., Temizel, C., 2018. Status of data-driven methods and their applications in oil and gas industry, in: Society of Petroleum Engineers - SPE Europec Featured at 80th EAGE Conference and Exhibition 2018. https://doi.org/10.2118/190812-ms

Barbosa, L.F.F.M., Nascimento, A., Mathias, M.H., de Carvalho, J.A., 2019. Machine learning methods applied to drilling rate of penetration prediction and optimization - A review. J. Pet. Sci. Eng. 183, 106332. https://doi.org/10.1016/j.petrol.2019.106332

Barnard, E., Wessels, L.F.A., 1992. Extrapolation and Interpolation in Neural Network Classifiers. IEEE Control Syst. 12. https://doi.org/10.1109/37.158898

Belazreg, L., Mahmood, S.M., 2020. Water alternating gas incremental recovery factor prediction and WAG pilot lessons learned. J. Pet. Explor. Prod. Technol. https://doi.org/10.1007/s13202-019-0694-x

Belazreg, L., Mahmood, S.M., Aulia, A., 2020. Random forest algorithm for co2 water alternating gas incremental recovery factor prediction. Int. J. Adv. Sci. Technol.

Belazreg, L., Mahmood, S.M., Aulia, A., 2019. Novel approach for predicting water alternating gas injection recovery factor. J. Pet. Explor. Prod. Technol. https://doi.org/10.1007/s13202-019-0673-2

Bian, X.Q., Han, B., Du, Z.M., Jaubert, J.N., Li, M.J., 2016. Integrating support vector regression with genetic algorithm for CO2-oil minimum miscibility pressure (MMP) in pure and impure CO2 streams. Fuel. https://doi.org/10.1016/j.fuel.2016.05.124

Bikmukhametov, T., Jäschke, J., 2020. Combining machine learning and process engineering physics towards enhanced accuracy and explainability of data-driven models. Comput. Chem. Eng. https://doi.org/10.1016/j.compchemeng.2020.106834

Blanes de Oliveira, L.A., de Carvalho Carneiro, C., 2021. Synthetic geochemical well logs generation using ensemble machine learning techniques for the Brazilian pre-salt reservoirs. J. Pet. Sci. Eng. 196, 108080. https://doi.org/10.1016/j.petrol.2020.108080

British Petroleum., 2021. Statistical Review of World Energy 2021. BP Energy Outlook 2021 70.

Busby, D., Pivot, F., Tadjer, A., 2017. Use of data analytics to improve well placement optimization under uncertainty, in: Society of Petroleum Engineers - SPE Abu Dhabi International Petroleum Exhibition and Conference 2017. https://doi.org/10.2118/188265-ms

Cao, Q., Banerjee, R., Gupta, S., Li, J., Zhou, W., Jeyachandra, B., 2016. Data driven production forecasting using machine learning, in: Society of Petroleum Engineers - SPE Argentina Exploration and Production of Unconventional Resources Symposium. https://doi.org/10.2118/180984-ms

Chaikine, I.A., Gates, I.D., 2021. A machine learning model for predicting multi-stage horizontal well production. J. Pet. Sci. Eng. https://doi.org/10.1016/j.petrol.2020.108133

Chaki, S., Zagayevskiy, Y., Shi, X., Wong, T., Noor, Z., 2020. Machine learning for proxy modeling of dynamic reservoir systems: Deep neural network DNN and recurrent neural network RNN applications, in: International Petroleum Technology Conference 2020, IPTC 2020. https://doi.org/10.2523/iptc-20118-ms

Chen, G., Fu, K., Liang, Z., Sema, T., Li, C., Tontiwachwuthikul, P., Idem, R., 2014. The genetic algorithm based back propagation neural network for MMP prediction in CO2-EOR process. Fuel. https://doi.org/10.1016/j.fuel.2014.02.034

Chen, G., Zhang, K., Zhang, L., Xue, X., Ji, D., Yao, C., Yao, J., Yang, Y., 2020. Global and Local Surrogate-Model-Assisted Differential Evolution for Waterflooding Production Optimization. SPE J. https://doi.org/10.2118/199357-PA

Chen, Y., Zhu, Z., Lu, Y., Hu, C., Gao, F., Li, W., Sun, N., Feng, T., 2019. Reservoir recovery estimation using data analytics and neural network based analogue study, in: Society of Petroleum Engineers - SPE/IATMI Asia Pacific Oil and Gas Conference and Exhibition 2019, APOG 2019. https://doi.org/10.2118/196487-ms

Chu, M. gon, Min, B., Kwon, S., Park, G., Kim, S., Huy, N.X., 2020. Determination of an infill well placement using a data-driven multi-modal convolutional neural network. J. Pet. Sci. Eng. https://doi.org/10.1016/j.petrol.2019.106805

Crombecq, K., 2011. Surrogate modeling of computer experiments with sequential experimental design. Ghent University.

Cross, T., Sathaye, K., Darnell, K., Niederhut, D., Crifasi, K., 2020. Predicting Water Production in the Williston Basin using a Machine Learning Model. https://doi.org/10.15530/urtec-2020-2756

da Silva, L.M., Avansi, G.D., Schiozer, D.J., 2020. Support vector regression for petroleum reservoir production forecast considering geostatistical realizations. SPE Reserv. Eval. Eng. https://doi.org/10.2118/203828-PA

Dachanuwattana, S., Jin, J., Zuloaga-Molero, P., Li, X., Xu, Y., Sepehrnoori, K., Yu, W., Miao, J., 2018. Application of proxy-based MCMC and EDFM to history match a Vaca Muerta shale oil well. Fuel. https://doi.org/10.1016/j.fuel.2018.02.018

Dai, Z., Middleton, R., Viswanathan, H., Fessenden-Rahn, J., Bauman, J., Pawar, R., Lee, S.-Y., McPherson, B., 2014. An integrated framework for optimizing CO2 sequestration and enhanced oil recovery. Environ. Sci. Technol. Lett. 1, 49–54.

Dargahi-Zarandi, A., Hemmati-Sarapardeh, A., Shateri, M., Menad, N.A., Ahmadi, M., 2020. Modeling minimum miscibility pressure of pure/impure CO2-crude oil systems using adaptive boosting support vector regression: Application to gas injection processes. J. Pet. Sci. Eng. https://doi.org/10.1016/j.petrol.2019.106499

Deng, L., Pan, Y., 2020. Machine-learning-assisted closed-loop reservoir management using echo state network for mature fields under waterflood. SPE Reserv. Eval. Eng. https://doi.org/10.2118/200862-PA

Dige, N., Diwekar, U., 2018. Efficient sampling algorithm for large-scale optimization under uncertainty problems. Comput. Chem. Eng. 115, 431–454. https://doi.org/10.1016/j.compchemeng.2018.05.007

Ding, S., Lu, R., Xi, Y., Liu, G., Ma, J., 2020. Efficient well placement optimization coupling hybrid objective function with particle swarm optimization algorithm. Appl. Soft Comput. 95, 106511.

Dong, P., Liao, X., Chen, Z., Chu, H., 2019. An improved method for predicting CO2 minimum miscibility pressure based on artificial neural network. Adv. Geo-Energy Res. https://doi.org/10.26804/ager.2019.04.02

Ehsan, A., Leung, J.Y., Zanon, S.D., Dzurman, P.J., 2014. An integrated application of cluster analysis and artificial neural networks for SAGD recovery performance prediction in heterogeneous reservoirs, in: Society of Petroleum Engineers - SPE Heavy Oil Conference Canada 2014. https://doi.org/10.2118/170113-ms

Enab, K., Ertekin, T., 2020. Screening and optimization of CO2-WAG injection and fish-bone well structures in low permeability reservoirs using artificial neural network. J. Pet. Sci. Eng. https://doi.org/10.1016/j.petrol.2020.108268

Enick, R.M., Olsen, D.K., Ammer, J.R., Schuller, W., others, 2012. Mobility and Conformance Control for CO2 EOR via Thickeners, Foams, and Gels--A Literature Review of 40 Years of Research and Pilot Tests, in: SPE Improved Oil Recovery Symposium.

Ersahin, A., Ertekin, T., 2020. Artificial Neural Network Modeling of Cyclic Steam Injection Process in

Naturally Fractured Reservoirs, in: SPE Reservoir Evaluation and Engineering. https://doi.org/10.2118/195307-PA

Ertekin, T., Sun, Q., 2019. Artificial intelligence applications in reservoir engineering: A status check. Energies. https://doi.org/10.3390/en12152897

Ezugwu, A.E., Adeleke, O.J., Akinyelu, A.A., Viriri, S., 2020. A conceptual comparison of several metaheuristic algorithms on continuous optimisation problems. Neural Comput. Appl. 32. https://doi.org/10.1007/s00521-019-04132-w

Forrester, A.I.J., Sóbester, A., Keane, A.J., 2008. Engineering design via surrogate modelling : a practical guide. J. Wiley.

Ghriga, M.A., Grassl, B., Gareche, M., Khodja, M., Lebouachera, S.E.I., Andreu, N., Drouiche, N., 2019. Review of recent advances in polyethylenimine crosslinked polymer gels used for conformance control applications. Polym. Bull. 1–29.

Gogna, A., Tayal, A., 2013. Metaheuristics: Review and application. J. Exp. Theor. Artif. Intell. 25. https://doi.org/10.1080/0952813X.2013.782347

Guo, Z., Reynolds, A.C., 2018. Robust life-cycle production optimization with a support-vector-regression proxy. SPE J. https://doi.org/10.2118/191378-PA

Haley, P.J., Soloway, D., 2003. Extrapolation limitations of multilayer feedforward neural networks. https://doi.org/10.1109/ijcnn.1992.227294

Halton, J.H., 1960. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. Numer. Math. 2. https://doi.org/10.1007/BF01386213

Hammersley, J.M., Handscomb, D.C., 1964. Monte Carlo Methods, Monte Carlo Methods. https://doi.org/10.1007/978-94-009-5819-7

Han, B., Bian, X., 2018. A hybrid PSO-SVM-based model for determination of oil recovery factor in the low-permeability reservoir. Petroleum. https://doi.org/10.1016/j.petlm.2017.06.001

Hanga, K.M., Kovalchuk, Y., 2019. Machine learning and multi-agent systems in oil and gas industry applications: A survey. Comput. Sci. Rev. https://doi.org/10.1016/j.cosrev.2019.08.002

Hassan, A., Elkatatny, S., Abdulraheem, A., 2019. Application of artificial intelligence techniques to predict the well productivity of fishbone wells. Sustain. https://doi.org/10.3390/su11216083

Hassani, H., Sarkheil, H., 2011. A proxy modeling approach to optimization horizontal well placement. 45th US Rock Mech. ….

Hassani, H., Sarkheil, H., Foroud, T., Karimpooli, S., 2011. A proxy modeling approach to optimization horizontal well placement. 45th US Rock Mech. / Geomech. Symp.

He, Q., Mohaghegh, S.D., Liu, Z., 2016. Reservoir simulation using smart proxy in SACROC unit - Case study, in: SPE Eastern Regional Meeting. https://doi.org/10.2118/184069-MS

Hemmati-Sarapardeh, A., Amar, M.N., Soltanian, M.R., Dai, Z., Zhang, X., 2020a. Modeling CO2 Solubility in Water at High Pressure and Temperature Conditions. Energy and Fuels. https://doi.org/10.1021/acs.energyfuels.0c00114

Hemmati-Sarapardeh, A., Larestani, A., Nait Amar, M., Hajirezaie, S., 2020b. Applications of Artificial Intelligence Techniques in the Petroleum Industry. Gulf Professional Publishing.

Honorio, J., Chen, C., Gao, G., Du, K., Jaakkola, T., 2015. Integration of PCA with a novel machine learning method for reparameterization and assisted history matching geologically complex reservoirs, in: Proceedings - SPE Annual Technical Conference and Exhibition. https://doi.org/10.2118/175038-ms

Hourfar, F., Bidgoly, H.J., Moshiri, B., Salahshoor, K., Elkamel, A., 2019. A reinforcement learning approach for waterflooding optimization in petroleum reservoirs. Eng. Appl. Artif. Intell. https://doi.org/10.1016/j.engappai.2018.09.019

Huang, Z., Chen, Z., 2021. Comparison of different machine learning algorithms for predicting the SAGD production performance. J. Pet. Sci. Eng. 202, 108559. https://doi.org/10.1016/j.petrol.2021.108559

Illarionov, E., Temirchev, P., Voloskov, D., Gubanova, A., Koroteev, D., Simonov, M., Akhmetov, A., Margarit, A., 2020. 3D reservoir model history matching based on machine learning technology, in: Society of Petroleum Engineers - SPE Russian Petroleum Technology Conference 2020, RPTC 2020. https://doi.org/10.2118/201924-ru

Jaber, A.K., Al-Jawad, S.N., Alhuraishawy, A.K., 2019a. A review of proxy modeling applications in numerical reservoir simulation. Arab. J. Geosci. https://doi.org/10.1007/s12517-019-4891-1

Jaber, A.K., Alhuraishawy, A.K., AL-Bazzaz, W.H., 2019b. A data-driven model for rapid evaluation of miscible CO2-WAg flooding in heterogeneous clastic reservoirs, in: Society of Petroleum Engineers - SPE Kuwait Oil and Gas Show and Conference 2019, KOGS 2019. https://doi.org/10.2118/198013-ms

Jang, I., Oh, S., Kim, Y., Park, C., Kang, H., 2018. Well-placement optimisation using sequential artificial neural networks. Energy Explor. Exploit. https://doi.org/10.1177/0144598717729490

Jesmani, M., Jafarpour, B., Bellout, M.C., Foss, B., 2020. A reduced random sampling strategy for fast robust well placement optimization. J. Pet. Sci. Eng. 184, 106414.

Jia, D., Liu, H., Zhang, J., Gong, B., Pei, X., Wang, Q., Yang, Q., 2020. Data-driven optimization for fine water injection in a mature oil field. Pet. Explor. Dev. https://doi.org/10.1016/S1876-3804(20)60084-2

Jia, H., Deng, L., 2018. Water flooding flowing area identification for oil reservoirs based on the method of streamline clustering artificial intelligence. Pet. Explor. Dev. https://doi.org/10.1016/S1876-3804(18)30036-3

Kalam, S., Abu-Khamsin, S.A., Al-Yousef, H.Y., Gajbhiye, R., 2020. A novel empirical correlation for waterflooding performance prediction in stratified reservoirs using artificial intelligence. Neural Comput. Appl. https://doi.org/10.1007/s00521-020-05158-1

Karacan, C.Ö., 2020. A fuzzy logic approach for estimating recovery factors of miscible CO2-EOR projects in the United States. J. Pet. Sci. Eng. https://doi.org/10.1016/j.petrol.2019.106533

Karimi, M., Mortazavi, A., Ahmadi, M., 2017. Applying an optimized proxy-based workflow for fast history matching. Arab. J. Geosci. https://doi.org/10.1007/s12517-017-3247-y

Karkevandi-Talkhooncheh, A., Rostami, A., Hemmati-Sarapardeh, A., Ahmadi, M., Husein, M.M., Dabir, B., 2018. Modeling minimum miscibility pressure during pure and impure CO2 flooding using hybrid of radial basis function neural network and evolutionary techniques. Fuel. https://doi.org/10.1016/j.fuel.2018.01.101

Khan, M.R., Alnuaim, S., Tariq, Z., Abdulraheem, A., 2019. Machine learning application for oil rate prediction in artificial gas lift wells, in: SPE Middle East Oil and Gas Show and Conference, MEOS, Proceedings. https://doi.org/10.2118/194713-ms

Khazali, N., Sharifi, M., Ahmadi, M.A., 2019. Application of fuzzy decision tree in EOR screening assessment. J. Pet. Sci. Eng. https://doi.org/10.1016/j.petrol.2019.02.001

Kim, S., Lee, K., Lim, J., Jeong, H., Min, B., 2020. Development of ensemble smoother–neural network and its application to history matching of channelized reservoirs. J. Pet. Sci. Eng. https://doi.org/10.1016/j.petrol.2020.107159

Kristoffersen, B.S., Silva, T., Bellout, M., Berg, C.F., 2020. An automatic well planner for efficient well placement optimization under geological uncertainty, in: ECMOR 2020 - 17th European Conference on the Mathematics of Oil Recovery. https://doi.org/10.3997/2214-4609.202035211

Li, H., Yu, H., Cao, N., Tian, H., Cheng, S., 2020. Applications of Artificial Intelligence in Oil and Gas Development. Arch. Comput. Methods Eng. https://doi.org/10.1007/s11831-020-09402-8

Liu, Y., Durlofsky, L.J., 2020. Multilevel Strategies and Geological Parameterizations for History Matching Complex Reservoir Models, in: SPE Journal. https://doi.org/10.2118/193895-PA

Ma, H., Yu, G., She, Y., Gu, Y., 2019. Waterflooding optimization under geological uncertainties by using deep reinforcement learning algorithms, in: Proceedings - SPE Annual Technical Conference and Exhibition. https://doi.org/10.2118/196190-ms

Ma, Z., Leung, J.Y., 2020. Design of warm solvent injection processes for heterogeneous heavy oil reservoirs: A hybrid workflow of multi-objective optimization and proxy models. J. Pet. Sci. Eng. https://doi.org/10.1016/j.petrol.2020.107186

Mahmoud, A.A., Elkatatny, S., Al-AbdulJabbar, A., 2021. Application of machine learning models for real-time prediction of the formation lithology and tops from the drilling parameters. J. Pet. Sci. Eng. 203, 108574. https://doi.org/10.1016/j.petrol.2021.108574

Masini, S.R., Goswami, S., Kumar, A., Chennakrishnan, B., 2019. Decline curve analysis using artificial intelligence, in: Society of Petroleum Engineers - Abu Dhabi International Petroleum Exhibition and Conference 2019, ADIP 2019. https://doi.org/10.2118/197932-ms

Masoudi, R., Mohaghegh, S.D., Yingling, D., Ansari, A., Amat, H., Mohamad, N., Sabzabadi, A., Mandel, D., 2020. Subsurface analytics case study: Reservoir simulation and modeling of highly complex offshore field in Malaysia, using artificial intelligent and machine learning, in: Proceedings - SPE Annual Technical Conference and Exhibition. https://doi.org/10.2118/201693-ms

McKay, M.D., Beckman, R.J., Conover, W.J., 1979. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. Technometrics. https://doi.org/10.2307/1268522

Mesbah, M., Shahsavari, S., Soroush, E., Rahaei, N., Rezakazemi, M., 2018. Accurate prediction of miscibility of CO2 and supercritical CO2 in ionic liquids using machine learning. J. CO2 Util. https://doi.org/10.1016/j.jcou.2018.03.004

Mohaghegh, S., 2018. Data-Driven Analytics for the Geological Storage of CO2. CRC Press, Boca Raton, Florida.

Mohaghegh, S.D., 2022. Smart Proxy Modeling: Artificial Intelligence and Machine Learning in Numerical Simulation. CRC Press, Boca Raton. https://doi.org/10.1201/9781003242581

Mohaghegh, S.D., 2020. Subsurface analytics: Contribution of artificial intelligence and machine learning to reservoir engineering, reservoir modeling, and reservoir management. Pet. Explor. Dev. https://doi.org/10.1016/S1876-3804(20)60041-6

Mohaghegh, S.D., 2017a. Data-Driven Reservoir Modeling, Society of Petroleum Engineers.

Mohaghegh, S.D., 2017b. Shale Analytics, Shale Analytics. https://doi.org/10.1007/978-3-319-48753-3

Mohaghegh, S.D., 2013. Reservoir modeling of shale formations. J. Nat. Gas Sci. Eng. https://doi.org/10.1016/j.jngse.2013.01.003

Mohaghegh, Shahab Dean, 2011. Reservoir simulation and modeling based on artificial intelligence and data mining (AI&DM). J. Nat. Gas Sci. Eng. https://doi.org/10.1016/j.jngse.2011.08.003

Mohaghegh, Shahab D., 2011. Reservoir simulation and modeling based on pattern recognition, in: Society of Petroleum Engineers - SPE Digital Energy Conference and Exhibition 2011. https://doi.org/10.2118/143179-ms

Mohaghegh, S.D., 2005. Recent developments in application of artificial intelligence in petroleum engineering. JPT, J. Pet. Technol. 57. https://doi.org/10.2118/89033-JPT

Mohaghegh, S.D., Amini, S., Gholami, V., Gaskari, R., Bromhal, G.S., 2012a. Grid-Based Surrogate Reservoir Modeling (SRM) for Fast Track Analysis of Numerical Reservoir Simulation Models at the Gridblock Level, in: SPE Western Regional Meeting. Society of Petroleum Engineers. https://doi.org/10.2118/153844-MS

Mohaghegh, S.D., Hafez, H., Gaskari, R., Haajizadeh, M., Kenawy, M., 2006. Uncertainty analysis of a giant oil field in the middle east using surrogate reservoir model, in: 12th Abu Dhabi International Petroleum Exhibition and Conference, ADIPEC 2006: Meeting the Increasing Oil and Gas Demand Through Innovation. https://doi.org/10.2523/101474-ms

Mohaghegh, S.D., Liu, J., Gaskari, R., Maysami, M., Olukoko, O., 2012b. Application of well-based Surrogate Reservoir Models (SRMs) to two offshore fields in Saudi Arabia, case study, in: Society of Petroleum Engineers Western Regional Meeting 2012. https://doi.org/10.2118/153845-ms

Mohaghegh, S.D., Liu, J., Gaskari, R., Maysami, M., Olukoko, O.A., 2012c. Application of surrogate reservoir model (SRM) to an onshore green field in Saudi Arabia; case study, in: Society of Petroleum Engineers - North Africa Technical Conference and Exhibition 2012, NATC 2012: Managing Hydrocarbon Resources in a Changing Environment. https://doi.org/10.2118/151994-ms

Mohagheghian, E., James, L.A., Haynes, R.D., 2018. Optimization of hydrocarbon water alternating gas in the Norne field: Application of evolutionary algorithms. Fuel. https://doi.org/10.1016/j.fuel.2018.01.138

Mohammadi, K., Ameli, F., 2019. Toward mechanistic understanding of Fast SAGD process in naturally fractured heavy oil reservoirs: Application of response surface methodology and genetic algorithm. Fuel 253, 840–856.

Mohammadpoor, M., Torabi, F., 2020. Big Data analytics in oil and gas industry: An emerging trend. Petroleum. https://doi.org/10.1016/j.petlm.2018.11.001

Morales, A., Nasrabadi, H., Zhu, D., 2010. A modified genetic algorithm for horizontal well placement optimization in gas condensate reservoirs, in: Proceedings - SPE Annual Technical Conference and Exhibition. https://doi.org/10.2118/135182-ms

Morales, A.N., Nasrabadi, H., Zhu, D., 2011. A new modified genetic algorithm for well placement optimization under geological uncertainties, in: 73rd European Association of Geoscientists and Engineers Conference and Exhibition 2011: Unconventional Resources and the Role of Technology. Incorporating SPE EUROPEC 2011. https://doi.org/10.2118/143617-ms

Mousavi, S.M., Jabbari, H., Darab, M., Nourani, M., Sadeghnejad, S., 2020. Optimal well placement using machine learning methods: Multiple reservoir scenarios, in: Society of Petroleum Engineers - SPE Norway Subsurface Conference 2020. https://doi.org/10.2118/200752-ms

Na-udom, A., Rungrattanaubol, J., 2015. A Comparison of Artificial Neural Network and Kriging Model for Predicting the Deterministic Output Response. NU. Int. J. Sci. 10, 1–9.

Nait Amar, M., Ghriga, M.A., Ouaer, H., El Amine Ben Seghier, M., Pham, B.T., Andersen, P.Ø., 2020a. Modeling viscosity of CO2 at high temperature and pressure conditions. J. Nat. Gas Sci. Eng. https://doi.org/10.1016/j.jngse.2020.103271

Nait Amar, M., Hemmati-Sarapardeh, A., Varamesh, A., Shamshirband, S., 2019. Predicting solubility of CO2 in brine by advanced machine learning systems: Application to carbon capture and sequestration. J. CO2 Util. https://doi.org/10.1016/j.jcou.2019.05.009

Nait Amar, M., Jahanbani Ghahfarokhi, A., 2020. Prediction of CO2 diffusivity in brine using white-box machine learning. J. Pet. Sci. Eng. https://doi.org/10.1016/j.petrol.2020.107037

Nait Amar, M., Jahanbani Ghahfarokhi, A., Ng, C.S.W., Zeraibi, N., 2021. Optimization of WAG in real geological field using rigorous soft computing techniques and nature-inspired algorithms. J. Pet. Sci. Eng. 109038. https://doi.org/10.1016/j.petrol.2021.109038

Nait Amar, M., Jahanbani Ghahfarokhi, A., Zeraibi, N., 2020b. Predicting thermal conductivity of carbon dioxide using group of data-driven models. J. Taiwan Inst. Chem. Eng. https://doi.org/10.1016/j.jtice.2020.08.001

Nait Amar, M., Zeraibi, N., 2019. An efficient methodology for multi-objective optimization of water alternating CO2 EOR process. J. Taiwan Inst. Chem. Eng. https://doi.org/10.1016/j.jtice.2019.03.016

Nait Amar, M., Zeraibi, N., 2018. Application of hybrid support vector regression artificial bee colony for prediction of MMP in CO2-EOR process. Petroleum. https://doi.org/10.1016/j.petlm.2018.08.001

Nait Amar, M., Zeraibi, N., Jahanbani Ghahfarokhi, A., 2020c. Applying hybrid support vector regression and genetic algorithm to water alternating CO2 gas EOR. Greenh. Gases Sci. Technol. https://doi.org/10.1002/ghg.1982

Nait Amar, M., Zeraibi, N., Redouane, K., 2018a. Optimization of WAG Process Using Dynamic Proxy, Genetic Algorithm and Ant Colony Optimization. Arab. J. Sci. Eng. https://doi.org/10.1007/s13369-018-3173-7

Nait Amar, M., Zeraibi, N., Redouane, K., 2018b. Bottom hole pressure estimation using hybridization neural networks and grey wolves optimization. Petroleum. https://doi.org/10.1016/j.petlm.2018.03.013

Nait Amar, M., Zeraibi, N., Redouane, K., 2018c. Pure CO2-Oil System Minimum Miscibility Pressure Prediction Using Optimized Artificial Neural Network by Differential Evolution. Pet. Coal 60.

Negash, B.M., Yaw, A.D., 2020. Artificial neural network based production forecasting for a hydrocarbon reservoir under water injection. Pet. Explor. Dev. https://doi.org/10.1016/S1876-3804(20)60055-6

Ng, C.S.W., Ghahfarokhi, A.J., Nait Amar, M., 2022a. Production optimization under waterflooding with Long Short-Term Memory and metaheuristic algorithm. Petroleum. https://doi.org/10.1016/J.PETLM.2021.12.008

Ng, C.S.W., Jahanbani Ghahfarokhi, A., Nait Amar, M., 2022b. Well production forecast in Volve field: Application of rigorous machine learning techniques and metaheuristic algorithm. J. Pet. Sci. Eng. 208, 109468. https://doi.org/10.1016/J.PETROL.2021.109468

Ng, C.S.W., Jahanbani Ghahfarokhi, A., Nait Amar, M., 2021a. Well Production Forecast in Volve field: Application of Machine Learning Techniques and Metaheuristic Algorithm . Pet. Sci. Technol.

Ng, C.S.W., Jahanbani Ghahfarokhi, A., Nait Amar, M., 2021b. Application of nature-inspired algorithms and artificial neural network in waterflooding well control optimization. J. Pet. Explor. Prod. Technol. https://doi.org/10.1007/s13202-021-01199-x

Ng, C.S.W., Jahanbani Ghahfarokhi, A., Nait Amar, M., Torsæter, O., 2021c. Smart Proxy Modeling of a Fractured Reservoir Model for Production Optimization: Implementation of Metaheuristic Algorithm and Probabilistic Application. Nat. Resour. Res. 30, 2431–2462. https://doi.org/10.1007/s11053-021-09844-2

Nwachukwu, A., Jeong, H., Pyrcz, M., Lake, L.W., 2018a. Fast evaluation of well placements in heterogeneous reservoir models using machine learning. J. Pet. Sci. Eng. https://doi.org/10.1016/j.petrol.2018.01.019

Nwachukwu, A., Jeong, H., Sun, A., Pyrcz, M., Lake, L.W., 2018b. Machine learning-based optimization of well locations and WAG parameters under geologic uncertainty, in: Proceedings - SPE Symposium on Improved Oil Recovery. https://doi.org/10.2118/190239-ms

Omrani, P.S., Vecchia, A.L., Dobrovolschi, I., van Baalen, T., Poort, J., Octaviano, R., Binn-Tahir, H., Muñoz, E., 2019. Deep learning and hybrid approaches applied to production forecasting, in: Society of Petroleum Engineers - Abu Dhabi International Petroleum Exhibition and Conference 2019, ADIP 2019. https://doi.org/10.2118/197498-ms

Onwunalu, J.E., Litvak, M.L., Durlofsky, L.J., Aziz, K., 2008. Application of Statistical Proxies to Speed Up Field Development Optimization Procedures, in: Abu Dhabi International Petroleum Exhibition and Conference. Society of Petroleum Engineers. https://doi.org/10.2118/117323-MS

Osarogiagbon, A.U., Oloruntobi, O., Khan, F., Venkatesan, R., Butt, S., 2020. Gamma ray log generation from drilling parameters using deep learning. J. Pet. Sci. Eng. 195, 107906. https://doi.org/10.1016/j.petrol.2020.107906

Otchere, D.A., Arbi Ganat, T.O., Gholami, R., Ridha, S., 2021. Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties: Comparative analysis of ANN and SVM models. J. Pet. Sci. Eng. https://doi.org/10.1016/j.petrol.2020.108182

Panja, P., Velasco, R., Pathak, M., Deo, M., 2018. Application of artificial intelligence to forecast hydrocarbon production from shales. Petroleum. https://doi.org/10.1016/j.petlm.2017.11.003

Panjalizadeh, H., Alizadeh, A., Ghazanfari, M., Alizadeh, N., 2015. Optimization of the WAG injection process. Pet. Sci. Technol. 33, 294–301.

Parada, C.H., Ertekin, T., 2012. A new screening tool for improved oil recovery methods using artificial neural networks, in: Society of Petroleum Engineers Western Regional Meeting 2012. https://doi.org/10.2118/153321-ms

Plaksina, T., 2019. Modern Data Analytics: Applied AI and Machine Learning for Oil and Gas Industry. Library and archives of Canada .

Pouladi, B., Keshavarz, S., Sharifi, M., Ahmadi, M.A., 2017. A robust proxy for production well placement optimization problems. Fuel. https://doi.org/10.1016/j.fuel.2017.06.030

Rahmanifard, H., Alimohammadi, H., Gates, I., 2020. Well Performance Prediction in Montney Formation Using Machine Learning Approaches. https://doi.org/10.15530/urtec-2020-2465

Rahmanifard, H., Plaksina, T., 2019. Application of artificial intelligence techniques in the petroleum industry: a review. Artif. Intell. Rev. https://doi.org/10.1007/s10462-018-9612-8

Rammay, M.H., Elsheikh, A.H., Chen, Y., 2020. Robust Algorithms for History Matching of Imperfect Subsurface Models. SPE J. https://doi.org/10.2118/193838-pa

Rana, S., Ertekin, T., King, G.R., 2018. An efficient assisted history matching and uncertainty quantification workflow using Gaussian processes proxy models and variogram based sensitivity analysis: GP-VARS. Comput. Geosci. https://doi.org/10.1016/j.cageo.2018.01.019

Redouane, K., Zeraibi, N., Nait Amar, M., 2019. Adaptive surrogate modeling with evolutionary algorithm for well placement optimization in fractured reservoirs. Appl. Soft Comput. J. https://doi.org/10.1016/j.asoc.2019.03.022

Rezaian, A., Kordestany, A., Haghighat, S.M., 2010. Experimental and artificial neural network approaches to predict the effect of PVA (Poly Vinyl Acetate) on the rheological properties of water and crude oil in EOR processes, in: Society of Petroleum Engineers - Nigeria Annual International Conference and Exhibition 2010, NAICE. https://doi.org/10.2118/140680-ms

Riazi, S.H., Zargar, G., Baharimoghadam, M., Moslemi, B., Sharifi Darani, E., 2016. Fractured reservoir history matching improved based on artificial intelligent. Petroleum. https://doi.org/10.1016/j.petlm.2016.09.001

Ross, T.J., 2010. Fuzzy Logic with Engineering Applications: Third Edition, Fuzzy Logic with Engineering Applications: Third Edition. https://doi.org/10.1002/9781119994374

Russell, S., Norvig, P., 2010. Artificial Intelligence A Modern Approach Third Edition, Pearson. https://doi.org/10.1017/S0269888900007724

Sampaio, T.P., Filho, V.J.M.F., De Sa Neto, A., 2009. An application of feed forward neural network as nonlinear proxies for the use during the history matching phase, in: SPE Latin American and Caribbean Petroleum Engineering Conference Proceedings. https://doi.org/10.2118/122148-ms

Sayyafzadeh, M., 2015a. A self-adaptive surrogate-assisted evolutionary algorithm for well placement optimization problems, in: Society of Petroleum Engineers - SPE/IATMI Asia Pacific Oil and Gas Conference and Exhibition, APOGCE 2015. https://doi.org/10.2118/176468-ms

Sayyafzadeh, M., 2015b. History Matching by Online Metamodeling, in: SPE Reservoir Characterisation and Simulation Conference and Exhibition. Society of Petroleum Engineers. https://doi.org/10.2118/175618-MS

Schulze-Riegert, R.W., Axmann, J.K., Haase, O., Rian, D.T., You, Y.L., 2002. Evolutionary algorithms applied to history matching of complex reservoirs. SPE Reserv. Eval. Eng. https://doi.org/10.2118/77301-PA

Shahkarami, A., Mohaghegh, S., 2020. Applications of smart proxies for subsurface modeling. Pet. Explor. Dev. https://doi.org/10.1016/S1876-3804(20)60057-X

Shahkarami, A., Mohaghegh, S., Gholami, V., Haghighat, A., Moreno, D., 2014a. Modeling pressure and saturation distribution in a CO2 storage project using a Surrogate Reservoir Model (SRM). Greenh. Gases Sci. Technol. https://doi.org/10.1002/ghg.1414

Shahkarami, A., Mohaghegh, S.D., Gholami, V., Haghighat, S.A., 2014b. Artificial intelligence (AI) assisted history matching, in: Society of Petroleum Engineers - SPE Western North American and Rocky Mountain Joint Meeting. https://doi.org/10.2118/169507-ms

Shahkarami, A., Mohaghegh, S.D., Hajizadeh, Y., 2018. Assisted history matching using pattern recognition technology. Int. J. Oil, Gas Coal Technol. https://doi.org/10.1504/IJOGCT.2018.090966

Shams, M., El-Banbi, A., Sayyouh, H., 2019. A Novel Assisted History Matching Workflow and its Application in a Full Field Reservoir Simulation Model. J. Pet. Sci. Technol. 9, 64–87.

Shannon, C.E., 1948. A Mathematical Theory of Communication. Bell Syst. Tech. J. 27. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Siena, M., Guadagnini, A., Della Rossa, E., Lamberti, A., Masserano, F., Rotondi, M., 2016. A novel enhanced-oil-recovery screening approach based on Bayesian clustering and principal-component analysis, in: SPE Reservoir Evaluation and Engineering. https://doi.org/10.2118/174315-PA

Sinha, S., de Lima, R.P., Lin, Y., Sun, A.Y., Symon, N., Pawar, R., Guthrie, G., 2020. Leak detection in carbon sequestration projects using machine learning methods: Cranfield site, Mississippi, USA, in: Proceedings - SPE Annual Technical Conference and Exhibition. https://doi.org/10.2118/201552-ms

Sinha, U., Dindoruk, B., Soliman, M., 2020. Prediction of CO2 minimum miscibility pressure MMP using machine learning techniques, in: Proceedings - SPE Symposium on Improved Oil Recovery. https://doi.org/10.2118/200326-ms

Sipöcz, N., Tobiesen, F.A., Assadi, M., 2011. The use of Artificial Neural Network models for CO2 capture plants. Appl. Energy. https://doi.org/10.1016/j.apenergy.2011.01.013

Sobol', I.M., 1967. On the distribution of points in a cube and the approximate evaluation of integrals. USSR Comput. Math. Math. Phys. https://doi.org/10.1016/0041-5553(67)90144-9

Sun, J., Ma, X., Kazi, M., 2018. Comparison of decline curve analysis DCA with recursive neural networks RNN for production forecast of multiple wells, in: SPE Western Regional Meeting Proceedings. https://doi.org/10.2118/190104-ms

Sun, Q., Ertekin, T., 2020. Screening and optimization of polymer flooding projects using artificial-neural-network (ANN) based proxies. J. Pet. Sci. Eng. https://doi.org/10.1016/j.petrol.2019.106617

Tatar, A., Shokrollahi, A., Mesbah, M., Rashid, S., Arabloo, M., Bahadori, A., 2013. Implementing Radial Basis Function Networks for modeling CO2-reservoir oil minimum miscibility pressure. J. Nat. Gas Sci. Eng. https://doi.org/10.1016/j.jngse.2013.09.008

Temizel, C., Aktas, S., Kirmaci, H., Susuz, O., Zhu, Y., Ranjith, R., Tahir, S., 2016. Turning data into knowledge: Data-driven surveillance and optimization in mature fields, in: Proceedings - SPE Annual Technical Conference and Exhibition. https://doi.org/10.2118/181881-ms

Thanh, H.V., Binh, D. Van, Kantoush, S.A., Nourani, V., Saber, M., Lee, K.-K., Sumi, T., 2022. Reconstructing Daily Discharge in a Megadelta Using Machine Learning Techniques. Water Resour. Res. 58, e2021WR031048. https://doi.org/https://doi.org/10.1029/2021WR031048

Tian, C., Horne, R.N., 2017. Recurrent neural networks for permanent downhole gauge data analysis, in: Proceedings - SPE Annual Technical Conference and Exhibition. https://doi.org/10.2118/187181-ms

Tilahun, S.L., 2019. Balancing the Degree of Exploration and Exploitation of Swarm Intelligence Using Parallel Computing. Int. J. Artif. Intell. Tools 28. https://doi.org/10.1142/S0218213019500143

Tillerson, R.W., 2008. Meeting global energy supply and demand challenges, in: 19th World Petroleum Congress.

Tunkiel, A.T., Sui, D., Wiktorski, T., 2020. Data-driven sensitivity analysis of complex machine learning models: A case study of directional drilling. J. Pet. Sci. Eng. 195, 107630. https://doi.org/10.1016/j.petrol.2020.107630

Urban-Rascon, E., Aguilera, R., 2020. Machine learning applied to SRV modeling, fracture characterization, well interference and production forecasting in low permeability reservoirs, in: SPE Latin American and Caribbean Petroleum Engineering Conference Proceedings. https://doi.org/10.2118/199082-ms

Vahdanikia, N., Divandari, H., Hemmati-Sarapardeh, A., Nait Amar, M., Schaffie, M., Ranjbar, M., 2020. Integrating new emerging technologies for enhanced oil recovery: Ultrasonic, microorganism, and emulsion. J. Pet. Sci. Eng. 192, 107229. https://doi.org/10.1016/j.petrol.2020.107229

Viana, F.A.C., 2016. A Tutorial on Latin Hypercube Design of Experiments. Qual. Reliab. Eng. Int. 32, 1975–1985. https://doi.org/10.1002/qre.1924

Vida, G., Shahab, M.D., Mohammad, M., 2019. Smart proxy modeling of SACROC CO2-EOR. Fluids. https://doi.org/10.3390/fluids4020085

Vo Thanh, H., Sugai, Y., Sasaki, K., 2020. Application of artificial neural network for predicting the performance of CO2 enhanced oil recovery and storage in residual oil zones. Sci. Rep. https://doi.org/10.1038/s41598-020-73931-2

Wang, L., Li, Z.P., Adenutsi, C.D., Zhang, L., Lai, F.P., Wang, K.J., 2021. A novel multi-objective optimization method for well control parameters based on PSO-LSSVR proxy model and NSGA-II algorithm. J. Pet.

Sci. Eng. https://doi.org/10.1016/j.petrol.2020.107694

Wei, W., Rezazadeh, A., Wang, J., Gates, I.D., 2021. An analysis of toe-to-heel air injection for heavy oil production using machine learning. J. Pet. Sci. Eng. 197, 108109. https://doi.org/10.1016/j.petrol.2020.108109

Wong, W.K., Ming, C.I., 2019. A Review on Metaheuristic Algorithms: Recent Trends, Benchmarking and Applications, in: 2019 7th International Conference on Smart Computing and Communications, ICSCC 2019. https://doi.org/10.1109/ICSCC.2019.8843624

Xiong, X., Lee, K.J., 2020. Data-driven modeling to optimize the injection well placement for waterflooding in heterogeneous reservoirs applying artificial neural networks and reducing observation cost. Energy Explor. Exploit. https://doi.org/10.1177/0144598720927470

Xu, K., Li, J., Zhang, M., Du, S.S., Kawarabayashi, K.I., Jegelka, S., 2020. How neural networks extrapolate: From feedforward to graph neural networks. arXiv.

Xu, T., 1998. Coupled modeling of non-isothermal multiphase flow, solute transport and reactive chemistry in porous and fractured media: 2. Model Applications.

Yang, L., Shami, A., 2020. On hyperparameter optimization of machine learning algorithms: Theory and practice. Neurocomputing 415, 295–316. https://doi.org/10.1016/J.NEUCOM.2020.07.061

Yang, T., Arief, I.H., Niemann, M., Houbiers, M., Meisingset, K.K., Martins, A., Froelich, L., 2019. A machine learning approach to predict gas oil ratio based on advanced mud gas data, in: Society of Petroleum Engineers - SPE Europec Featured at 81st EAGE Conference and Exhibition 2019. https://doi.org/10.2118/195459-ms

Yang, X.S., Chien, S.F., Ting, T.O., 2014. Computational intelligence and metaheuristic algorithms with applications. Sci. World J. https://doi.org/10.1155/2014/425853

Yazdanpanah, A., Hashemi, A., 2012. Production Optimization Using an Experimental Design and Genetic Algorithm. J. Am. Sci. 8.

Yeten, B., Castellini, A., Guyaguler, B., Chen, W.H., 2005. A Comparison Study on Experimental Design and Response Surface Methodologies, in: SPE Reservoir Simulation Symposium. Society of Petroleum Engineers. https://doi.org/10.2118/93347-MS

You, J., Ampomah, W., Kutsienyo, E.J., Sun, Q., Balch, R.S., Aggrey, W.N., Cather, M., 2019a. Assessment of enhanced oil recovery and CO2 storage capacity using machine learning and optimization framework, in: Society of Petroleum Engineers - SPE Europec Featured at 81st EAGE Conference and Exhibition 2019. https://doi.org/10.2118/195490-ms

You, J., Ampomah, W., Sun, Q., 2020a. Development and application of a machine learning based multi-objective optimization workflow for CO2-EOR projects. Fuel 264, 116758.

You, J., Ampomah, W., Sun, Q., 2020b. Co-optimizing water-alternating-carbon dioxide injection projects using a machine learning assisted computational framework. Appl. Energy. https://doi.org/10.1016/j.apenergy.2020.115695

You, J., Ampomah, W., Sun, Q., Kutsienyo, E.J., Balch, R.S., Cather, M., 2019b. Multi-objective optimization of CO2 enhanced oil recovery projects using a hybrid artificial intelligence approach, in: Proceedings - SPE Annual Technical Conference and Exhibition. https://doi.org/10.2118/196182-ms

You, J., Ampomah, W., Sun, Q., Kutsienyo, E.J., Balch, R.S., Dai, Z., Cather, M., Zhang, X., 2020c. Machine learning based co-optimization of carbon dioxide sequestration and oil recovery in CO2-EOR project. J. Clean. Prod. https://doi.org/10.1016/j.jclepro.2020.120866

Yousef, A.M., Kavousi, G.P., Alnuaimi, M., Alatrach, Y., 2020. Predictive data analytics application for enhanced oil recovery in a mature field in the Middle East. Pet. Explor. Dev. 47. https://doi.org/10.1016/S1876-3804(20)60056-8

Zarei, F., Daliri, A., Alizadeh, N., 2008. The use of Neuro-Fuzzy proxy in well placement optimization, in: Intelligent Energy Conference and Exhibition. Society of Petroleum Engineers. https://doi.org/10.2118/112214-MS

Zerafat, M.M., Ayatollahi, S., Mehranbod, N., Barzegari, D., 2011. Bayesian network analysis as a tool for efficient EOR screening, in: Society of Petroleum Engineers - SPE Enhanced Oil Recovery Conference 2011, EORC 2011. https://doi.org/10.2118/143282-ms

Zhong, Z., Sun, A.Y., Wang, Y., Ren, B., 2020. Predicting field production rates for waterflooding using a machine learning-based proxy model. J. Pet. Sci. Eng. https://doi.org/10.1016/j.petrol.2020.107574

Zubarev, D.I., 2009. Pros and cons of applying proxy-models as a substitute for full reservoir simulations, in: Proceedings - SPE Annual Technical Conference and Exhibition. https://doi.org/10.2118/124815-ms