

Article

Practical CO₂—WAG Field Operational Designs Using Hybrid Numerical-Machine-Learning Approaches

Qian Sun ^{1,2}, William Ampomah ^{1,*}, Junyu You ^{1,3}, Martha Cather ¹ and Robert Balch ¹

¹ Petroleum Recovery Research Center, New Mexico Institute of Mining and Technology, Socorro, NM 87801, USA; sunqian.psu@vip.126.com (Q.S.); junyu.you@student.nmt.edu (J.Y.); martha.cather@nmt.edu (M.C.); robert.balch@nmt.edu (R.B.)

² School of Energy Resources, China University of Geoscience, Beijing 100083, China

³ School of Petroleum and Natural Gas Engineering, Chongqing University of Science and Technology, Chongqing 401331, China

* Correspondence: william.ampomah@nmt.edu

Abstract: Machine-learning technologies have exhibited robust competences in solving many petroleum engineering problems. The accurate predictivity and fast computational speed enable a large volume of time-consuming engineering processes such as history-matching and field development optimization. The Southwest Regional Partnership on Carbon Sequestration (SWP) project desires rigorous history-matching and multi-objective optimization processes, which fits the superiorities of the machine-learning approaches. Although the machine-learning proxy models are trained and validated before imposing to solve practical problems, the error margin would essentially introduce uncertainties to the results. In this paper, a hybrid numerical machine-learning workflow solving various optimization problems is presented. By coupling the expert machine-learning proxies with a global optimizer, the workflow successfully solves the history-matching and CO₂ water alternative gas (WAG) design problem with low computational overheads. The history-matching work considers the heterogeneities of multiphase relative characteristics, and the CO₂-WAG injection design takes multiple techno-economic objective functions into accounts. This work trained an expert response surface, a support vector machine, and a multi-layer neural network as proxy models to effectively learn the high-dimensional nonlinear data structure. The proposed workflow suggests revisiting the high-fidelity numerical simulator for validation purposes. The experience gained from this work would provide valuable guiding insights to similar CO₂ enhanced oil recovery (EOR) projects.

Keywords: multi-objective optimization; CO₂-WAG; machine learning; numerical modeling; hybrid workflows



Citation: Sun, Q.; Ampomah, W.; You, J.; Cather, M.; Balch, R. Practical CO₂—WAG Field Operational Designs Using Hybrid Numerical-Machine-Learning Approaches. *Energies* **2021**, *14*, 1055. <https://doi.org/10.3390/en14041055>

Academic Editor: Valentina Colla

Received: 21 January 2021

Accepted: 12 February 2021

Published: 17 February 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The Southwest Regional Partnership on Carbon Sequestration (SWP) project focuses on the design and monitoring of a field-scale CO₂ enhanced oil recovery (EOR) process in the Farnsworth Unit (FWU) located in the Anadarko Basin, Texas. From 2010 to 2014, 16.82 billion standard cubic feet of anthropogenic CO₂ was injected into the Morrow-B sand [1]. The CO₂ utilized in this project is captured by the Arkalon Ethanol Plant and the Agrium Fertilizer Plant locating in Liberal, Kansas, and Borger, Texas, respectively [2]. According to Munson [3], the original oil (OOIP) and gas in place (OGIP) are approximately 120 million (MM) barrels and 41.48 billion standard cubic feet (SCF), respectively. The field development was initiated in 1955 and the waterflood started in 1963. Starting from the end of 2010, the FWU field has been undergoing a water alternative CO₂ injection process to extract the residual oil in place.

The Morrow B sandstone formation is located at a depth interval between 7550 ft and 7950 ft. The formation has an average dip angle of <1° [4] and was deposited in the late Pennsylvanian by a fluvial system in an incised valley [5]. The average net pay

thickness is 22 ft. The initial reservoir and bubble point pressure are 2203 psi and 2059 psi, respectively [4]. The reservoir temperature at a depth of 7900 ft is measured to be 168°F. The reservoir has a mean porosity of 14.6% and a mean permeability of 58 md [6]. The west half of the field is considered highly permeable and porous when compared with the east half [3]. The reservoir was initially undersaturated with gas–oil solution ratio of 345 SCF/stock tank barrel (STB), and the oil saturation and formation volume factor were characterized as 69% and 1.192 res bbl/STB, respectively [6].

Reservoir engineers have successfully structured many different versions of numerical simulation models to investigate the fluid flow transportation dynamics in Morrow-B formation and monitor the long-term fate of the CO₂ plume [7]. A numerical compositional model coupled with geological, geophysical, and engineering data was reported in previous work [8]. The initial SWP's FWU geological static model was first presented in 2015 [9] and was employed by many related works [10–12]. Investigations on the hydraulic flow unit (HFU) was performed to characterize the heterogeneous petrophysical properties [13]. A facies model was developed to populate the petrophysical property distributions with assistance from the HFU. There are numerous history-matching studies imposed on reservoir models considering the field historical injection and production data in the primary, secondary, and tertiary recovery period [14]. The history-matched model can be used to assess various CO₂ water-alternative-gas (WAG) forecasting scenarios and more importantly, optimize the project design strategies.

The machine-learning technologies exhibit strong competences to solve a large spectrum of petroleum engineering problems, including sweet spot identification [15], history matching [16], fluid property characterization [17], and field development strategy optimization [18]. In the field of reservoir simulation, the machine-learning models comprehend the fluid transportation dynamics in porous media via learning the data structure presented by a knowledge base instead of solving the partial differential equations using numerical and analytical methods [19]. The training of the machine-learning model takes advantages of the knowledge base containing field and synthetic data samples. To validate the machine-learning model, rigorous blind testing applications must be imposed to investigate the generalization capability. The blind testing performances of the machine-learning models used in this work were less than 1%. Notably, the computational overhead using the machine-learning model could be reduced by several orders of magnitudes more than using the high-fidelity numerical simulator [20]. Literature surveys indicate that machine-learning models have been successfully developed to simulate production performance of gas condensate reservoirs [21], shale gas reservoirs [22], coalbed methane reservoirs [23], enhanced oil recovery processes [24], etc. Moreover, for certain field specified problems, the machine-learning models can also be trained utilizing seismic, well-log, and production data, which is competent to assess the production performance within the area covered by the seismic survey [25]. In this case, the machine-learning model can make reliable predictions without the presence of hydrodynamic (permeability, relative permeability, etc.) and petrophysical (fluid composition, pressure-volume-temperature (PVT) data, etc.) data. Moreover, recent research efforts make attempts to develop machine-learning models as an alternative to the equation of state and conduct flash calculations in the compositional simulations [26]. Such well-trained machine-learning models (also known as expert systems) can be employed as “proxies” of the high-fidelity numerical simulator to generate a large volume of realizations in the computational demanding processes.

The SWP project would demand the use of machine-learning methods to solve the history-matching and CO₂-WAG injection optimization problems. As a carbon capture, utilization, and sequestration (CCUS) project, the field operator considers not only the oil recovery, but also the volume of CO₂ sequestration, both of which determine the project economics from different perspectives. When more objective functions are included, the computational overhead of the optimization study would be even heavier. In this article, a robust machine-learning assisted multi-objective optimization workflow is presented. Expert proxy models, a global optimization algorithm, and numerical models are effectively

coupled in the workflows to generate fast and high-quality reservoir engineering analysis. The discussion of this paper starts from the description of the reservoir model. Then the machine-learning proxies employed by this work are summarized. Afterwards, the optimization algorithms and treatment of the multi-objective problem are presented. By coupling the machine-learning proxy and the optimization protocols, two workflows are structured to solve the history-matching and CO₂-WAG design optimization problems. Last but not least, we employ case studies by imposing the proposed workflow on the FWU field.

2. Reservoir Modeling

A reservoir model is the most fundamental element of the proposed workflow. The establishment of the numerical simulation model assembles geological, petrophysical, and field historical data. Reservoir engineers have structured many versions of numerical simulation models for the FWU field considering hydrodynamic, geo-mechanical, and geochemical mechanisms of the fluid transportation in Morrow-B formation. The field scale numerical models are validated via rigorous history-matching processes using the historical injection and production data and employed to build forecasting scenarios. In addition, sector models around the well 13–10A pattern are utilized to validate the optimization workflows.

2.1. Hydraulic Flow Unit

The geological model takes advantage of the petrophysical properties interpolated from the laboratory measurements using the hydraulic flow unit (HFU). Morrow-B sandstone exhibits strong heterogeneity divergences from the diagenetic processes, which leads to various pore structures, multiphase flow characteristics, and wettability. Moreover, rock and fluid property variations could evolve progressively after CO₂-WAG injection starts. The porosity–permeability relationship can be identified using the HFU to classify the sedimentologic and diagenetic heterogeneity of the Morrow-B formation.

Morrow B sandstone was first classified into five porosity facies and eight subfaces; and eight HFUs were identified to characterize hydrogeologic heterogeneities using the well log and petrophysical data collected from various scientific wells [13]. Followed up research works suggest lumping HFU 5 to 8 into an identical group due to the similar flow features observed from core-flooding experiments [27].

The geological model continuously utilizes the permeability–porosity relationship developed from the HFUs. When a reservoir simulation model is structured, grids are assigned with different relative permeability curves based on the HFU characteristics. In Figure 1, the HFU distributions at different simulation layers of the reservoir model are displayed.

2.2. Updated Geological Model

As with the advancement of the reservoir characteristic analysis on the FWU field, the original simulation model is restructured using an updated outer boundary and property population protocol. As shown in Figure 2, the permeability and porosity distributions of the updated geological mode are displayed in Figure 2a,c using a 100 ft by 100 ft mesh grid system. To implement the updated geological model, the history-matching process must be revisited from the primary and secondary stages. Due to the long production period (from 1956 to 2010), completing one simulation case takes more than 12,000 s of central processing unit (CPU) time, which makes the computational cost of the history-matching process prohibitively expensive. With such a running speed, even preparing the dataset to train the proxy model becomes unrealistic. Thus, the upscale of the original model becomes necessary. Figure 2b,d show the upscaled permeability and porosity distribution using a 200 ft by 200 ft mesh grid system. Figure 3 illustrates the comparison of the simulation results regarding the oil and water production using the original and upscaled reservoir model with an average disparity of less than 0.8%. The use of the upscaled

model significantly reduces the computational time to less than 300 s of CPU time, which is 40 times faster than the original model. With the help of upscaling, the history-matching of the primary and secondary recovery periods is done using the coarser grid first and then validated by the finer grids.

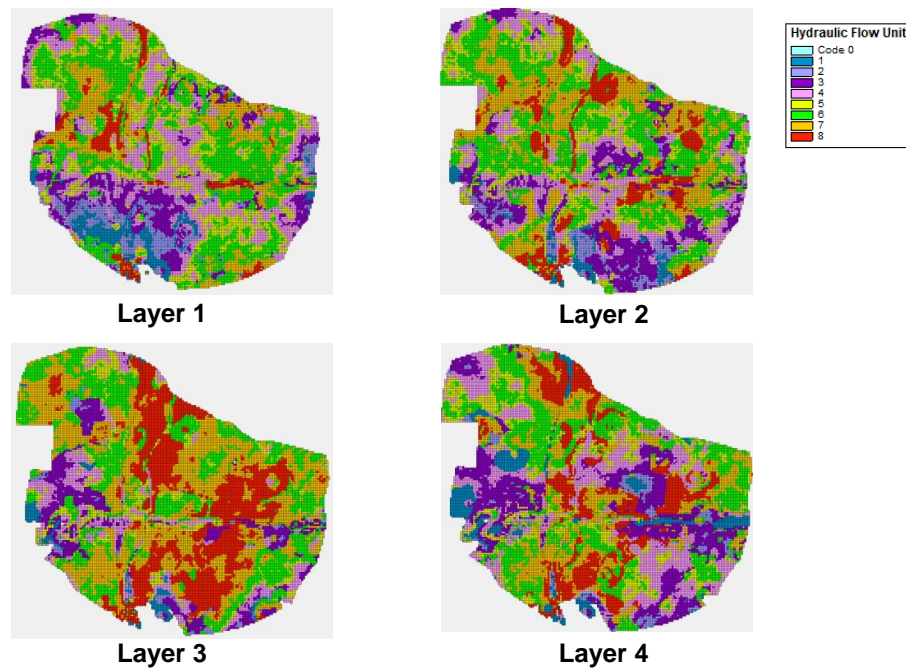


Figure 1. HFU (hydraulic flow unit) distributions in different layers of Morrow-B formation.

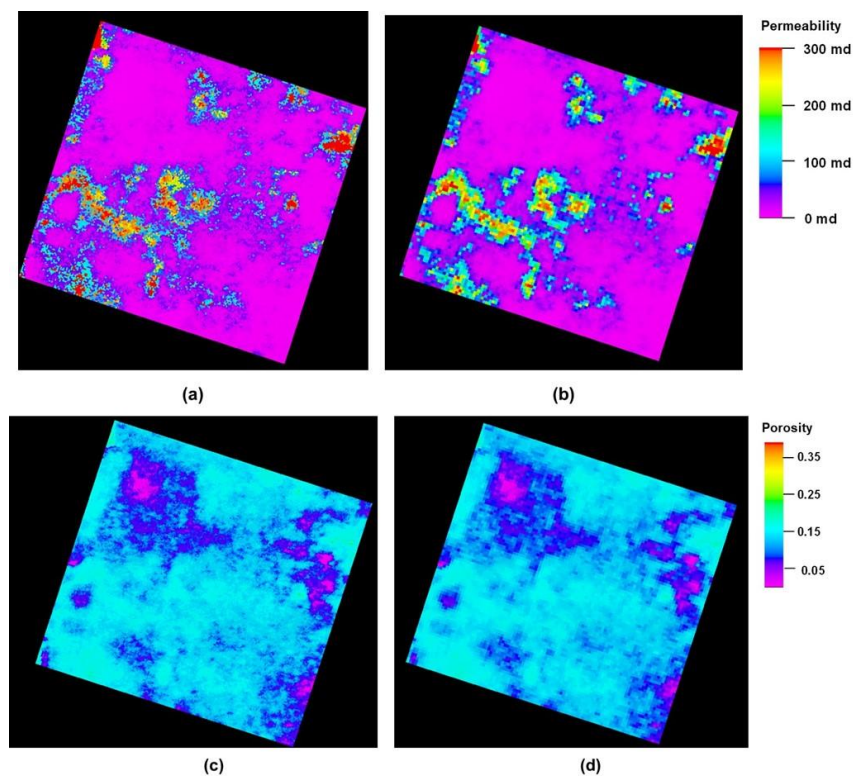


Figure 2. Updated geological model and the upscaled grid displayed as (a) permeability in a finer grid, (b) permeability in a coarse grid, (c) porosity in a finer grid, and (d) porosity in a coarse grid.

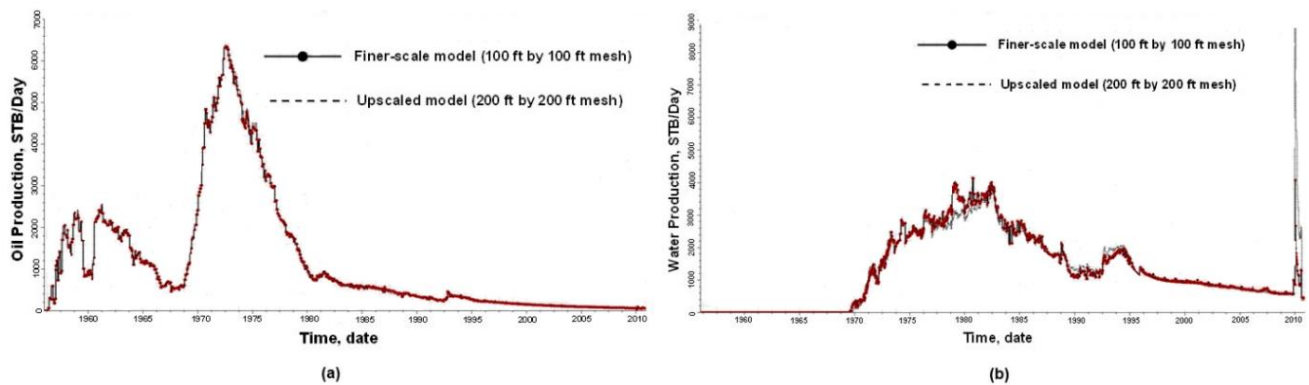


Figure 3. Comparison of the results generated by the finer and upscaled (coarser) grid simulation models in terms of (a) oil production and (b) water production.

2.3. Injection Pattern Model

A sector model focusing on the Well 13–10A pattern was structured to simulate the smaller area. As shown in Figure 4, the sector model is discretized using a 60 by 60 Cartesian grid system (mesh size is 26.7 ft by 26.7 ft) and simulates 1/8 of the five-spot injection pattern. The property distribution is established using the data collected from 13-A well. This work employs the sector model to generate datasets for the testing purposes of the optimization workflow, since the computational cost of completing one simulation is low. Moreover, the injection pattern model is suitable to investigate multiple mechanisms near the scientific well.

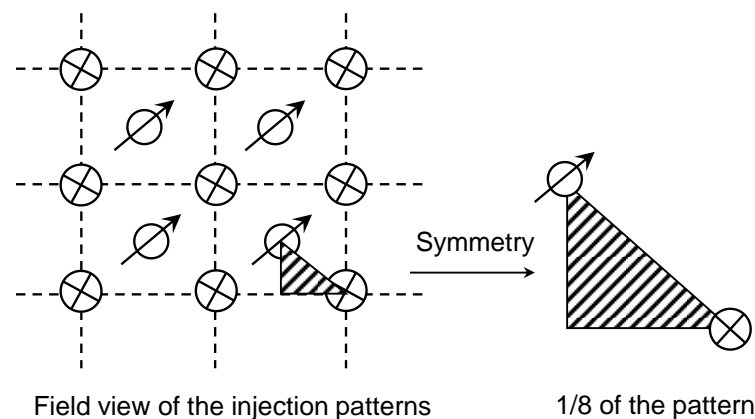


Figure 4. Illustration of the injection pattern-based model.

The aforementioned simulation models played a crucial role in this work in terms of comprehending the fluid flow dynamics in the Morrow-B sands, and more importantly, generating the desired data to develop the machine-learning proxy models. It is worth stressing that the interaction between the high-fidelity numerical and machine-learning protocols is critical for the long-term development and updating of the workflow.

3. Machine-Learning Proxies

In this work, the history-matching and CO₂ injection design processes employed many machine-learning technologies such as response surface models (RSM), multi-layer neural networks (MLNN), and support vector machines (SVM). These expert machine-learning models (also called “proxies”) are developed to learn from the numerical simulation models and investigate the fluid transportation mechanisms via a data-driven perspective. Due to their high computational efficacies and robust generalization competence, machine-

learning models have been successfully developed and act as regression tool to evaluate the field performance in terms of the fluid recovery, injection performance, pressure responses, and economic assessments.

In the history-matching applications, the proxy models correlate the uncertain hydro-dynamical properties with the fluid production/injection and pressure data, which assists the evaluation of the fitting error during the tuning process. There are many generations of reservoir simulation models evolving as more geological, petrophysical, and field operational data become available. The initial history matching works can be done without the help of machine-learning technologies because the earlier version of the reservoir model is simpler in terms of the reservoir characteristics. When the new geological model arrives, the permeability distributions are updated and more importantly, various relative permeability curves are assigned to the grids based on the HFU indices. Consequentially, the number of uncertain parameters becomes considerably larger, and so does the required volume of simulation realizations. Therefore, the use of machine-learning technologies becomes quite necessary to history-match the current version of the simulation model.

After a history-matched model was structured, another class of proxy model was developed to learn the data structure presented between the CO₂-WAG injection strategies and the forecasted project's techno-economic responses. The development of the proxy models includes the training process and a rigorous blind testing protocol to ensure the prediction accuracy. The following discussions summarize the machine-learning models developed in the history-matching and optimization studies of the SWP CO₂-WAG projects.

3.1. Response Surface Models

The response surface model (RSM) is one of the most classic nonlinear regression algorithms. The prediction model can be expressed via Equation (1):

$$y = b_0 + \sum_{i=1}^m b_i x_i + \sum_{i=1}^m \sum_{j \geq i}^m b_{ij} x_i x_j + \sum_{i=1}^m b_{ii} x_i^2 \quad (1)$$

where b_0 , b_i , b_{ii} , and b_{ij} are the regression coefficients; x_i is the i th input feature; m is the total number of input features; and y is the output feature. To avoid overfitting issues, the higher than quadratic order terms are usually omitted. Gradient decent and least square methods [28] can be used to determine the regression coefficient. The RSM model is the earliest machine-learning model used in the CO₂-WAG optimization study of the SWP project. Four RSM proxies are developed to predict oil and water productions, CO₂ in place, and production volumes in the CO₂-WAG injection period of the project [6]. The R^2 value observed by comparing the results predicted by proxy and high-fidelity numerical models is close to one with relative error values of less than 0.1%, which confirms the validity of the RSM proxies.

3.2. Multi-Layer Neural Networks

Multi-layer neural networks (MLNN) model a class of artificial neural network (ANN) model that is inspired by the signal transportation process of biological neural units. Typically, an MLNN model is composed of an input, an output, and several hidden layers. Artificial neurons are included in these layers to store, transform, and transport data signals. The numerical values connecting the neurons in adjacent layers are "weights." An MLNN model transfers the signal via Equation (2):

$$a_i^n = f_n \left(\sum_j w_{i,j}^n a_j^{n-1} + b_i^n \right) \quad (2)$$

where the notation n numbers the current layer ($n - 1$ is the previous layer) of the layer, a is the artificial neuron, and w and b are the weight and the bias, respectively. The index notations i and j number the layer and neuron, respectively. Notably, f_n is a transfer

function constraining the signal within the range of $[0, 1]$ or $[-1, 1]$, which enables the MLNN model to deal with the nonlinearity of the dataset. The sigmoid and linear functions are commonly imposed on the hidden and output layers, respectively. The training process of the MLNN model essentially minimizes the error function depicted by Equation (3):

$$E(w) = \frac{1}{2} \|o - t\|^2 \quad (3)$$

where w refers to the weight vector, and o and t are the prediction and training target, respectively. Thus, the training epoch of the MLNN iteratively updates w to minimize the error function until a prescribed stopping criterion is achieved. Many training algorithms such as the scaled conjugate gradient and the Levenberg–Marquardt and Bayesian regularization [29] methods are widely used in training MLNN models with complex topologies.

There are many MLNN applications in the SWP project: In the machine-learning-assisted history-matching process, MLNN models are trained to predict the field production responses with varying reservoir properties. Two representative blind test cases are shown in Figure 5. The proxy model is competent to predict the oil and water productions when the hydrodynamical properties are changed, with overall blind testing errors at 0.5% and 9.87%, respectively. As shown in Figure 5, the satisfactory blind testing performance indicates that the ANN models have been well trained and could be used to evaluate the history-matching error in the proposed workflow. Notably, such results have not been history-matched yet.

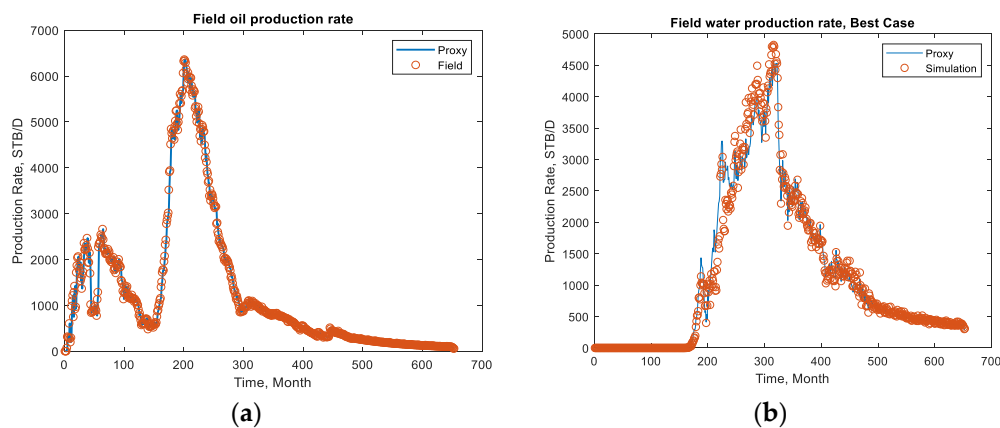


Figure 5. Blind testing performance of multi-layer neural networks (MLNN) proxies in the history-matching applications of (a) oil production and (b) water production.

The other class of MLNN models are trained to forecast the long-term field development responses in terms of hydrocarbon production, CO_2 sequestration volume, and project economic assessments considering the CO_2 -WAG design parameters as input [18]. The CO_2 -WAG operational parameters include the durations of CO_2 and water injection cycles, the water injection rate, and production well specifications. In this work, the error statistics of blind testing applications for an injection pattern and the field case applications, with overall mean errors observed as 0.71% and 1.00%, respectively. Therefore, the MLNN model can be employed as a surrogate model to evaluate the techno-economic objective functions considered in the design of CO_2 -WAG projects.

In the SWP project, MLNN models have been successfully trained and act as robust regression tools. The high computational speed enables many time-consuming studies such as global sensitivity analysis, history-matching, multi-objective optimization, etc.

3.3. Support Vector Machines

The support vector machine (SVM) is a robust regression model employed in this work. A linear SVM regression model is an extension of a linear regression model that aims at [30]:

$$\min \left(\frac{1}{2} \|\omega\|^2 \right) \quad (4)$$

$$\text{subject to } \begin{cases} y - \langle \omega, x \rangle - b \leq \varepsilon \\ \langle \omega, x \rangle + b - y \leq \varepsilon \end{cases} \quad (5)$$

The model is suitable to make predictions for problems with n input and 1 output, where $x \in \mathcal{R}^n$ and $y \in \mathcal{R}$. Note that in Equations (4) and (5), ω is the coefficient vector and b is the intercept. The flatness of the model measures how small the norm of ω could be. ε is the error tolerance of the regression problem. The operator $\langle x, \omega \rangle$ indicates the dot product of vectors x and ω . To extend the applicability of SVM model to solve real problems, the soft margin is introduced to Equation (6) via a slack variable ζ :

$$\min \left[\frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^m (\zeta_i + \zeta_i^*) \right] \quad (6)$$

$$\text{subject to } \begin{cases} y_i - \langle \omega, x_i \rangle - b \leq \varepsilon + \zeta_i \\ \langle \omega, x_i \rangle + b - y_i \leq \varepsilon + \zeta_i^* \\ \zeta_i, \zeta_i^* \geq 0, i = 1, \dots, m \end{cases}$$

where C is a constant to balance the flatness and the tolerance degree of the deviations larger than ε . A loss function depicted in Equation (7) is used:

$$|\zeta|_\varepsilon = \begin{cases} 0 & \text{if } |\zeta| \leq \varepsilon \\ |\zeta| - \varepsilon & \text{otherwise} \end{cases} \quad (7)$$

The modern SVM models are typically written via a dual optimization form, which can be expressed as follows:

$$\text{maximize } \begin{cases} -\frac{1}{2} \sum_{i,j=1}^m (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) \langle x_i, x_j \rangle \\ -\varepsilon \sum_{i=1}^m (\alpha_i + \alpha_i^*) + \sum_{i=1}^m y_i (\alpha_i - \alpha_i^*) \end{cases} \quad (8)$$

$$\text{subject to } \sum_{i=1}^m (\alpha_i - \alpha_i^*) = 0; \alpha_i, \alpha_i^* \in [0, C]$$

where α and α^* are Lagrange multipliers.

For nonlinear problems, kernel functions are employed to map the original training patterns into an implicit feature space so that the linear SVM regression can be used. In this case, Equation (8) is modified as Equation (9):

$$\text{maximize } \begin{cases} -\frac{1}{2} \sum_{i,j=1}^m (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) k(x_i, x_j) \\ -\varepsilon \sum_{i=1}^m (\alpha_i + \alpha_i^*) + \sum_{i=1}^m y_i (\alpha_i - \alpha_i^*) \end{cases} \quad (9)$$

where $k(x_i, x_j)$ is a kernel. In this work the Gaussian kernel written in Equation (10) was employed:

$$k(x_i, x_j) = \exp \left(-\frac{\|x_i - x_j\|^2}{2\sigma^2} \right) \quad (10)$$

Similar to the MLNN model, the training of an SVM model needs to identify three hyperparameters, including the constant C , magnitude of ϵ , and the kernel scale factor δ . In this work, the SVM models were developed to act as proxy models to predict the oil production (Proxy-Oil) and CO₂ storage volume (Proxy-CO₂). The cross validation results indicate that the relative error of the Proxy-Oil and Proxy-CO₂ comparing to the high-fidelity numerical simulator to be 0.06% and 0.01%, respectively [31].

By developing the machine-learning proxies, the following experiences can be summarized:

1. The RSM is more suitable for problems with a smaller input dimension and single output parameter. Compared to MLNN and SVM models, the training overhead is much lighter.
2. The MLNN model exhibits robust generalization capability for problems with large input and output dimensions. However, the degree of freedom of the hyperparameter is more than that of RSM and SVM models. Thus, more computational costs are required to obtain an optimum model with optimum prediction performance.
3. The SVM model is more suitable for problems with strong nonlinearity and a large input dimension. The number of hyperparameters to be tuned is smaller than in the MLNN model. However, it cannot make a prediction for more than one output variable.

All of the machine-learning proxies play vital roles in the optimization study by enabling various computational expensive processes, which require giant amounts of assessments on the objective functions.

4. Optimization Protocols

In this work, the metaheuristic and stochastic optimization protocols such as particle swarm optimization and genetic algorithms were employed to optimize various technical and economic objective functions. Compared to the conventional gradient-based optimization methods, such optimization protocols are not constrained by the continuity and differentiability of the objective function. Therefore, they are more suitable for solving problems with complex and implicit objective functions. However, the global optimization technologies used in this work would demand the establishment of populations with a certain volume of samples. Thus, totally relying on the high-fidelity numerical simulators would make the computational overhead prohibitively heavy. This section briefly goes through the critical strategies of global technologies included in this work.

4.1. Objective Functions and Constraints

Economic Objective Functions

The project net present value (NPV) is a major economic consideration in the optimization study. A general definition of NPV can be depicted as Equation (11):

$$NPV = CAPEX + \sum_{i=1}^m \frac{(q_{oi} \times \text{Oil price}) \times (1 - \text{tax rate}) - C_i}{(1 + \text{interest rate})^i} \quad (11)$$

In Equation (11), CAPEX is the project capital expenditure, m is the total number of counted timesteps of the project, q_{oi} is the cumulative oil production of the i th timestep, and C_i is the operational cost. In this work, the NPV definition was modified to adapt to the CO₂-WAG field operation as Equation (12):

$$NPV = CAPEX + \sum_{i=1}^n \frac{(q_o \times \text{Oil price} + q_{co2,store} \times r_{co2,credit} - C_i)}{(1 + \text{interest rate})^i} \quad (12)$$

where the CAPEX = 10 million dollars (USD), oil price is 50 USD/bbl, interest rate is 5%, the CO₂ storage credit ($r_{co2,credit}$) is 45 USD/metric ton. The cost term C_i includes water

and CO₂ injection cost, CO₂ purchase rate, and produced water treatment cost, which can be expressed by Equation (13):

$$C_i = q_{w,inj} \times r_w - q_{w,pro} \times r_{w,pro} - q_{inj,co2} \times r_{co2} - q_{co2,p} \times r_{co2,p} \quad (13)$$

where the water injection cost (r_w) is 1.03 USD/STB, the produced water treatment cost ($r_{w,pro}$) is 0.64 USD/STB, the CO₂ purchase cost is 1.72 USD/thousand(M) SCF, CO₂ injection cost is 0.85 USD/MSCF. The variables $q_{w,inj}$ and $q_{w,pro}$ are water injection and production rates in STB/day, respectively; $q_{inj,CO2}$ and $q_{CO2,p}$ are CO₂ injection and purchasing rates in MSCF/day, respectively.

The evaluation of the NPV value requires the oil/CO₂/water production and CO₂/water injection data reported by the numerical simulation model, which is determined by the CO₂-WAG design strategies. When the NPV is considered one of the objective functions in the optimization workflow, a large volume of realizations is required to find the optimum CO₂-WAG injection protocol to maximize the NPV. In this work, the calculation of the NPV was done by the proxy models with a low computational cost.

4.2. Technical Objective Functions

History matching error: The history-matching error is a significant objective function to measure the misfit of the numerical model results with the field historical data. In this work, the square error was used to account for the differences between the results generated by the simulator with the field historical measurement for oil, water, gas production data, water, gas injection data, and pressure data via Equation (14) through Equation (19), respectively.

$$E_o = \sum_i^m [(O_o - T_o)_i]^2 \quad (14)$$

$$E_{wp} = \sum_i^m [(O_{wp} - T_{wp})_i]^2 \quad (15)$$

$$E_{gp} = \sum_i^m [(O_{gp} - T_{gp})_i]^2 \quad (16)$$

$$E_{wi} = \sum_i^m [(O_{wi} - T_{wi})_i]^2 \quad (17)$$

$$E_{gi} = \sum_i^m [(O_{gi} - T_{gi})_i]^2 \quad (18)$$

$$E_p = \sum_i^m [(O_p - T_p)_i]^2 \quad (19)$$

where O and T represent the model prediction and historical data, respectively, and m is the total number of timesteps of the available field data. Notably, in this work, O was generated by machine-learning proxy models to reduce the computational overhead.

CO₂ storage volume: For a CCUS project, the CO₂ storage volume is a critical technical objective function considered in the optimization study, which can be defined as [6]:

$$\text{CO}_2 \text{ storage} = \text{CO}_2 \text{ purchased} - \text{CO}_2 \text{ produced} + \text{Recycle} \quad (20)$$

The CO₂ storage volume plays a crucial role in the project economic perspective by bringing considerable tax allowance to the project. Since the SWP project has rigorous plans to strategically purchase a certain volume of CO₂ based on the surface operational facility capacities, maximizing the CO₂ storage volume is essentially minimizing the CO₂ production volume in the CO₂-WAG process.

Oil recovery: The cumulative oil production (which could be normalized by the residual oil in place as the recovery factor) in the forecasting period is another technical

objective function to be maximized in the design of the CO₂-WAG project, which also decides the major income of the project NPV.

Physical and Engineering Constraints

The optimization protocols design the CO₂-WAG operational criteria from the algorithmic perspective. Without imposing proper physical engineering and physical constraints, the optimization results would be unrealistic. In this work, the following constraints were imposed on the optimization workflow to ensure the engineering practicability:

1. The history-matching study specifies the oil production, CO₂, and water injection rates, and considers the CO₂ and water production the primary objective functions. The constraints imposed on the history-matching work is that the average pressure must be below 5400 psi.
2. The CO₂-WAG optimization is constrained by an average reservoir pressure range of [3700, 5400] psi to maintain the miscibility of the sweeping front.

In the optimization process, either for the history-matching or project design purposes, the solutions in the population pool would be screened by the constraints to sustain the engineering applicability. To meet such a requirement, proxy models need to be developed to predict the average reservoir pressure based on various history-matching and CO₂-WAG design scenarios.

4.3. Treatment of Multiple-Objective Optimizations

The optimization studies in this work considered more than one objective function, which are called multi-objective optimization problems (MOO). For instance, the history-matching work needs to minimize the error functions defined by Equation (14) through Equation (19) simultaneously, and the CO₂-WAG design problem aims at maximizing the NPV, oil recovery, and CO₂ storage volume in the meantime. There were two different ways employed in this work to treat the MOO problems:

Aggregated method: An aggregated objective function can be defined by Equation (21):

$$f_a = \sum_{i=1}^m (w_i \times f_i) \quad (21)$$

where f_i refers to various objective functions, w_i is the associated weights applied to the functions, and f_a is the aggregated objective function. In this way, the MOO is converted to a single-objective optimization study. In this work, the weighted method was used to solve the history matching problem by defining the history matching error as Equation (22):

$$f_{HME} = E_o + E_{gp} + E_{wp} + E_{wi} + E_{gi} \quad (22)$$

where the weights imposed on the error terms are identical. In addition, the optimization of the CO₂-WAG design uses the aggregated objective function depicted in Equation (23) [6]:

$$f = w_1 \times \text{Cumulative Oil Production} + w_2 \times \text{CO}_2 \text{ Storage} + w_3 \times \text{NPV} \quad (23)$$

The weight factors can be justified based on the operational preference.

Pareto optimum theory: An alternative approach to treating the MOO problem is to generate a Pareto front solution by employing the Pareto optimum theory, which is suitable for establishing a solution repository considering the tradeoff relationship amongst multiple objective functions [32].

The Pareto optimum theory defines a vector $\vec{x} = [x_1, x_2, x_3, \dots, x_n]$ with n variables, and a vector $\vec{f}(\vec{x}) = \left[\vec{f}_1(\vec{x}), \vec{f}_2(\vec{x}), \vec{f}_3(\vec{x}), \dots, \vec{f}_m(\vec{x}) \right]$ with m objective functions considered in the optimization process. Ω assembles the candidate solutions in the searching domain. A solution $\vec{u} = [u_1, u_2, u_3, \dots, u_m]$ is defined to dominate

solution $\vec{v} = [v_1, v_2, v_3, \dots, v_m]$, where \vec{u} and \vec{v} are two objective function vectors, if $\forall i \in \{1, 2, 3, \dots, m\}$ to make $u_i \leq v_i$ and $\exists i \in \{1, 2, 3, \dots, m\}$ to make $u_i < v_i$. This concept can be denoted as $\vec{u} \preceq \vec{v}$. The Pareto optimum set collects the solution vectors as $P^* = \{x \in \Omega | \exists x' \in \Omega \vec{f}(x') \preceq \vec{f}(x)\}$ and the corresponding objective function ensemble is defined as the Pareto Front: $PF^* = \left\{ \vec{f} = [f_1(\vec{x}), f_2(\vec{x}), f_3(\vec{x}), \dots, f_m(\vec{x})] \mid x \in P^* \right\}$.

The Pareto optimum theory can be illustrated by Figure 6, which is a hypothetical problem aiming at minimizing f_1 and f_2 . The solutions \vec{x}_1^* and \vec{x}_2^* are two solutions in the Pareto optimum set. \vec{x}_1^* intends to improve objective function f_2 compared to \vec{x}_2^* by sacrificing (increasing) f_1 due to the tradeoff relationship. In addition, \vec{x}' is a dominating solution since solution \vec{x}_1^* exists, which exhibits smaller f_1 and f_2 at the same time.

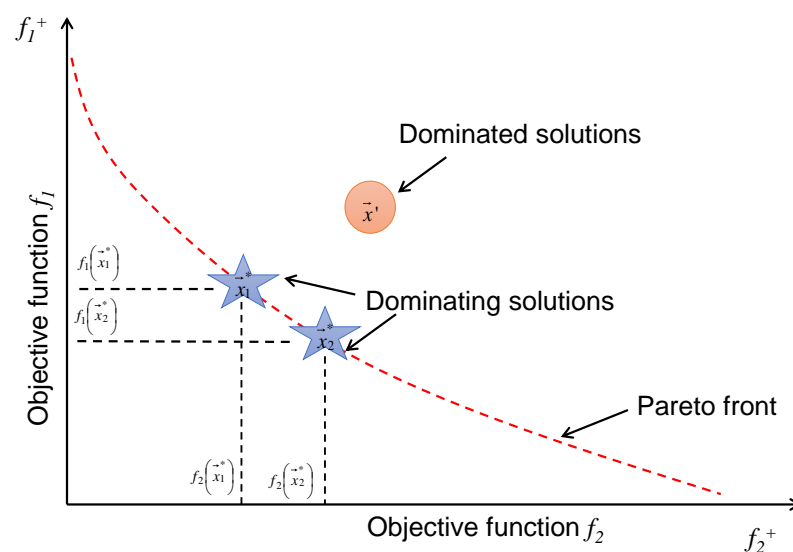


Figure 6. Pareto front, dominated, and dominating solutions.

It should be emphasized that regardless of the optimization algorithm, the application of the Pareto front theory by itself is a computational expensive process since it ranks a large volume of solutions regarding multiple objective functions. The CO₂-EOR project is a good candidate upon which to impose the Pareto front theory since the incremental oil production would yield more produced CO₂, which reduces the volume of carbon sequestration. More importantly, considering the project NPV, the impacts of the tax allowance and the oil production benefits need to be comprehensively investigated. More importantly, the Pareto front solution suggests the operators have more flexibility to design the CO₂-WAG processes under various techno-economic conditions. However, such an optimization protocol is not suitable for the history-matching problems. The reason is that history-matching processes shoot for a solution that minimizes all the objective functions simultaneously. A solution fitting the oil production history with large water production error could be a dominating solution in the Pareto front, but it cannot be considered in the history-matching study. Therefore, the majority of the Pareto optimum set would be abandoned.

4.4. Optimization Algorithms

Genetic algorithm (GA): The GA refers to a family of computational models derived from Darwin's theory of biological evolution. The idea is one of the natural selection organization principles for optimizing the individuals of populations. GAs mimic natural selection to optimize more successfully. Problems are solved by an evolutionary process resulting in a best solution (fittest survivor). GAs do not search via gradients; the searching

is done via sampling by stochastic operators rather than by deterministic rules. Figure 7 briefly illustrates the workflow of a GA optimizer. The application of a GA includes the following operators:

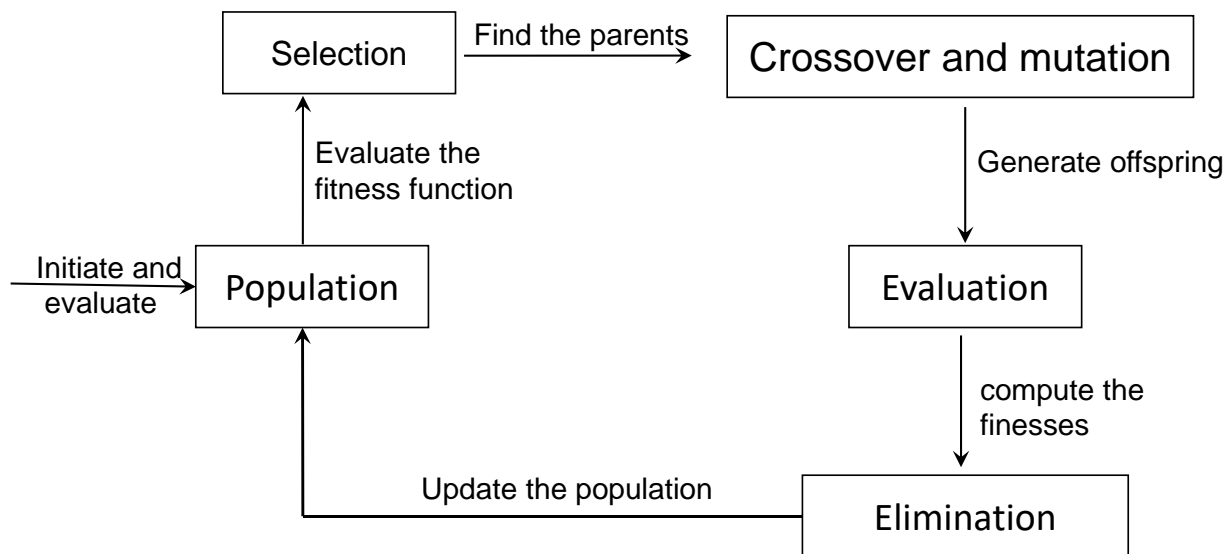


Figure 7. Optimization workflow using genetic algorithm (GA).

Encoding: The very first step is to convert the solution of a problem, typically composed of vectors of control parameters into a chromosome-type data structure.

Selection: The selection operator is employed to find individuals that exhibit good fitness functions and become candidates as parents to generate offspring. The most used selection operators include tournament and roulette wheel selection.

Reproducing: Once the parents are selected, the next step is to make them reproduce offspring via crossover. The crossover is a GA operator to generate new chromosomes based on the individuals of previous generations. A certain crossover rate needs to be specified so that a certain volume of selected individuals will be used to generate offspring. Typically, the crossover rate ranges from 60% to 95%.

Mutation: Mutation is a GA operator that makes the search range wider. What mutation does is change a gene of a chromosome completely without any reason. In a genetic algorithm, the mutation rate is extremely low, typically less than 1%.

Decoding: At the end of a genetic algorithm iteration, the chromosomes need to be converted back to real number solutions.

The GA is one of the first trials in the SWP project and optimizes multiple techno-economic objective functions for the CO₂-WAG injection process [6].

Particle swarm optimization (PSO): PSO is a nature-inspired evolutionary and stochastic optimization technique to solve computationally hard optimization problems. It is a fast technology based on the movement and intelligence of swarms. PSO was built by mimicking the working mechanisms of biological swarm migrations. In a PSO application, particles communicate directly or indirectly with one another via searching directions. During the iteration process, the particles update their position according to their previous experience and the experience of their neighbors. The process of PSO optimization includes the following three steps [33]:

1. Evaluate the fitness by the proxy model.
2. Calculate the velocity term v using Equation (24):

$$v_i(k+1) = wv_i(k) + c_1r_1[\bar{x}_i(k) - x_i(k)] + c_2r_2[g(k) - x_i(k)] \quad (24)$$

where k is the iteration level, \bar{x} is the local best, and g is the global best.

3. Update the particle position via Equation (25):

$$x_i(k+1) = x_i(k) + v_i(k+1) \quad (25)$$

Compared to GA, the optimization process of PSO does not include any encoding/decoding procedure, which accelerates the convergence. However, the PSO can be easily trapped by the local minima when a complex objective function is considered. Figure 8 is the optimization workflow using a PSO algorithm.

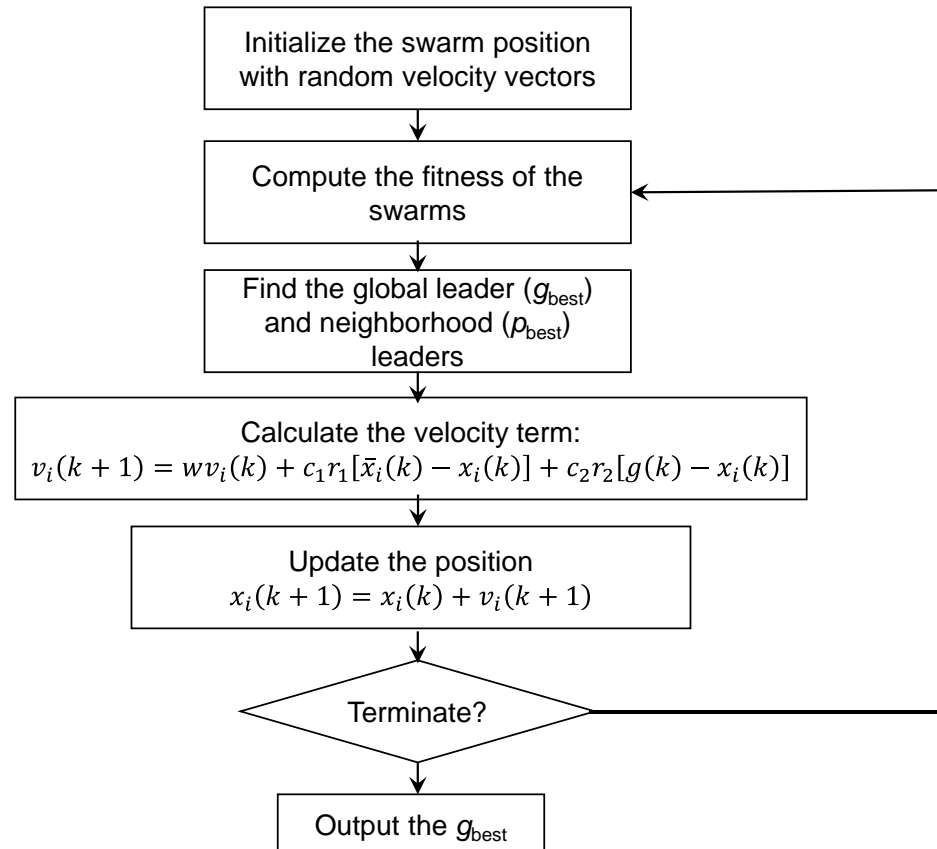


Figure 8. Optimization workflow using particle swarm optimization (PSO).

Multi-objective particle swarm optimization (MOPSO): Compared to the single-objective PSO, MOPSO introduces a “repository (REP)” to restore the nondominated solutions of each iteration. The size of the repository can be prescribed from 100 to 250 particles. When the repository is full, the dominated solution in REP needs to be replaced by the nondominated solutions appearing in the next generation.

A “hypercube” is introduced in MOPSO to quantify the “neighbor” (also known as the searching domain) of the particles. Hypercubes are a subdivision of the objective function domain separated by hyperlines that are uniformly distributed in each objective function domain. We illustrated a hypercube (Figure 9) of a problem with two objective functions of project NPV and oil recovery. The dots represent the particles of certain MOPSO iterations. It is worth emphasizing that the hyperlines need to be justified after the fitness of population (POP) is updated in each iteration due to the upper and lower limits of the objective function domains possibly being changed.

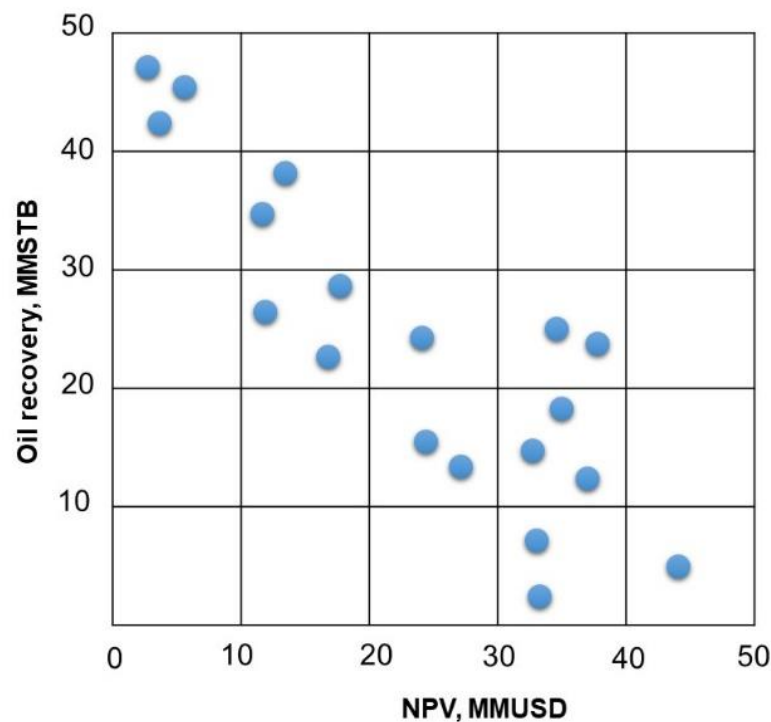


Figure 9. Hypercube illustration of a problem with two objective functions.

Similar to the single objective PSO, the MOPSO iteration also includes three stages [34]:

Stage 1: Initialization: The initial population (*POP*) is generated and the velocity (*VEL*) of each particle is set to 0. The objective functions of all the particles are computed and their fitness is assessed. Then, the initial solution repository (*REP*) is structured using the nondominated solutions in the *POP*. Then the initial personal best (*PBEST*) of the particles is set to be the initial fitness.

Stage 2: Velocity calculation: Compute the movement velocity of the particles using Equation (26):

$$VEL[i] = 0.4 \times VEL[i] + C_1(PBEST[i] - POP[i]) + C_2(REP[h] - POP[i]) \quad (26)$$

where C_1 and C_2 are two weight factors within $[0, 1]$. Based on the hypercube structure in the solution repository, a global best *REP* [*h*] is randomly picked from the grid where each particle is located.

Stage 3: Update: The particles in the population are updated via Equation (27):

$$POP[i] = POP[i] + VEL[i] \quad (27)$$

Then, the fitness of the particles is re-evaluated based on the updated population. At this stage, the *REP* needs to be updated by removing the dominating solutions. The hypercube structure and the *PBEST* would also change accordingly.

Current research works still focus on quantifying the convergence criteria of MOPSO. One of the most broadly accepted opinions to determine the convergence of MOPSO is that the iteration can be terminated when none of the new particles can dominate any of the particles involved in *REP*. Figure 10 shows the workflow of the MOPSO process. MOPSO is proved to be a robust algorithm that finds the repository of the close to the true Pareto front by many searches.

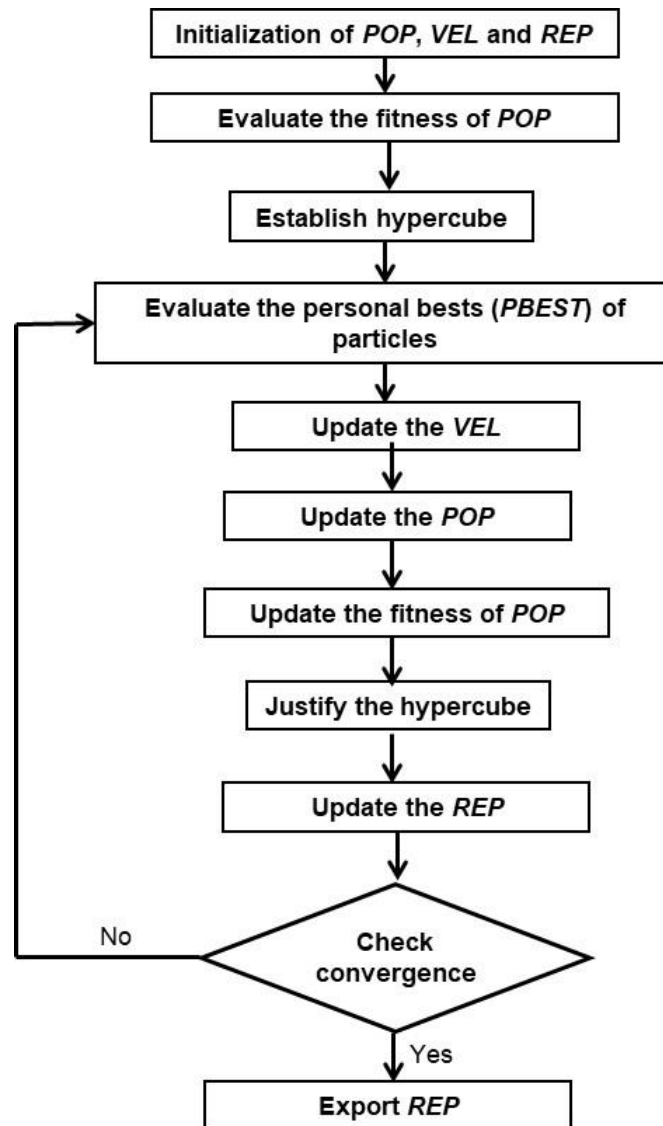


Figure 10. Workflow of multi-objective particle swarm optimization (MOPSO).

5. Structuring the Hybrid Numerical Machine-Learning Workflow

In the previous sections, we introduced the machine-learning proxy models and optimization protocols. The coupling of them would build robust machine-learning assisted optimization protocols for various engineering purposes in the CO₂-WAG injection project. In this paper, a machine-learning assisted history-matching and a multi-objective optimization workflow is presented.

5.1. History-Matching Workflow

Due to the high dimension and considerable computational overhead, history-matching cannot be completed by totally relying on the high-fidelity numerical model. Therefore, machine-learning models are needed to assist the history-matching work. Using the prepared numerical simulation runs as the knowledge base, a series of ANN models can be successfully trained to predict the oil, water, and gas production considering the uncertain reservoir properties as inputs. The importance of using the proxy models is that one can conduct a large volume of numerical simulation realizations with little computational cost. The trained proxy models are coupled with PSO to minimize the history-matching error

function. The error function is defined as the summation of the square differences between the field historical data and the model prediction.

In the history-matching process, the oil, gas, and water production rates were considered primary unknowns to calculate the error function. Considering the extensive number of tuning parameters, the results of the history-matching model could exhibit strong non-uniqueness, which means that there is more than one combination of the 62 parameters that could generate a similar level of history-matching error. After an optimal solution is obtained, the history matching solution found by the machine-learning workflow is revisited by the high-fidelity numerical reservoir simulator, which confirms the matching quality. In Figure 11, the machine-learning assisted history-matching workflow is displayed, which was successfully imposed to develop history-matched reservoir models for the primary/secondary and CO₂-WAG period.

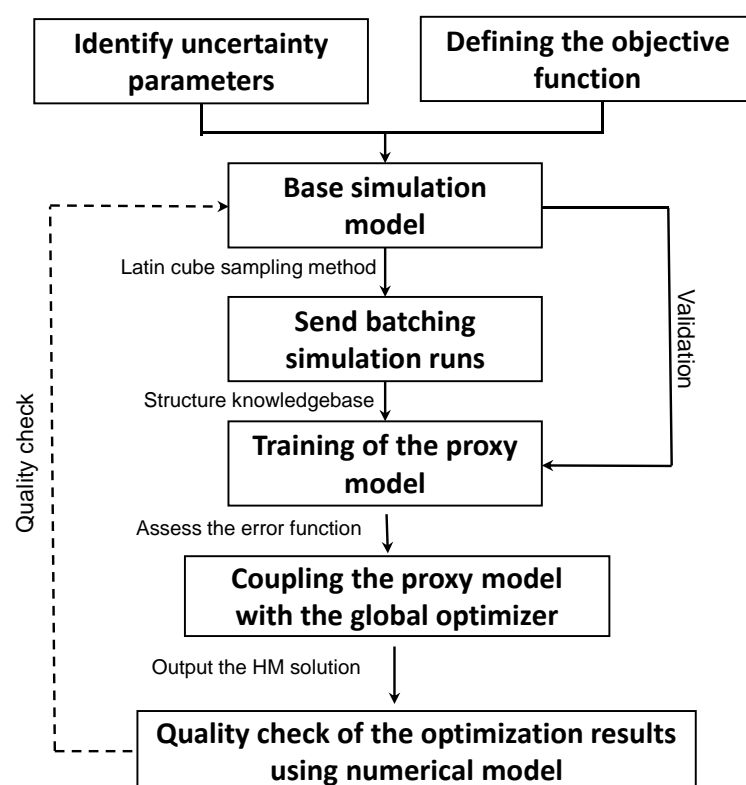


Figure 11. Machine-learning assisted history-matching workflow.

5.2. Multi-Objective Optimization Workflow

Figure 12 illustrate the machine-learning assisted multi-objective optimization workflow coupling the numerical reservoir simulation model, proxy models, and global optimization algorithm.

Data preparation: The original high-fidelity numerical model is utilized to prepare a certain volume of numerical realizations. Those realizations are utilized to prepare the knowledge base that is used to train the proxy model. Considering a blind testing error margin of <5%, the training of an injection pattern base model needs at least 100 simulation runs due to the smaller model size and dimensions. Training for the field scale proxy model may require more than 500 simulation runs.

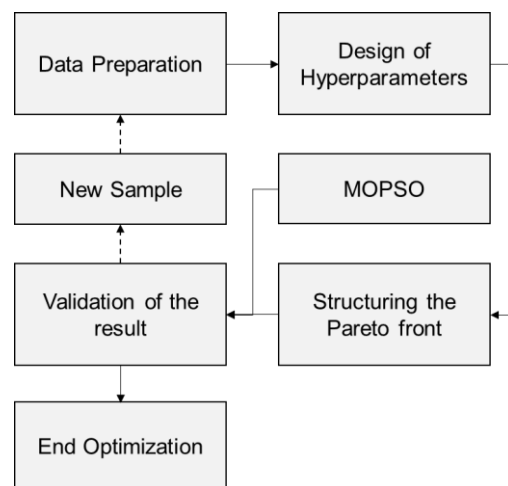


Figure 12. Machine-learning assisted multi-objective optimization workflow.

Design of hyperparameters: It is known that the hyperparameters of the machine-learning models have significant impacts on the prediction performance. The hyperparameters of SVM are optimized using Bayesian optimization and the MLNN topology is designed with a self-adaptive training protocol [35].

Structuring the Pareto front: MOPSO employs the proxy models to generate the Pareto front by considering various objective functions. The 2- and 3-objective Pareto fronts can be visualized using a two-dimensional curve and a surface, respectively.

Validation of the result: The Pareto front solutions must be validated before advising any operational decisions. The input parameters of the Pareto front solutions revisit the numerical model and compare the between the proxy and simulation results. If large disparities are observed, then some new samples are added to the training database to re-train the proxy. This loop can be continued several times until the error between the simulation results and the proxy predictions is lower than a prescribed error tolerance.

6. Case Studies

6.1. A History-Matching Application

A history-matching application using the proposed machine-learning assisted workflow is presented in this discussion. The objective of this study was to tune the reservoir hydrodynamic properties including the permeability along the x -, y -, and z - directions, and the Corey's relative permeability coefficients. The permeability distributions were tuned by imposing anisotropic multipliers. The reservoir model assigned various three-phase relative permeability curves based on the HFU characterization. There were five different relative permeability sets considered in this work. Notably, HFU 5, 6, 7, and 8 were lumped into one group and shared identical relative permeability data. Considering the permeability multiplier and five different sets of Corey's coefficients, there were 62 parameters involved in the history-matching processes.

The machine-learning assisted tuning process was carried out by the workflow discussed in the previous sections. The Latin cube sampling protocol was used to prepare 100 simulation cases by varying the uncertain hydrodynamic parameters. The dataset was used to train an expert MLNN proxy to predict the field responses based on different hydrodynamic properties. In the machine-learning assisted workflow, the proxy model substituted the high-fidelity numerical simulator to evaluate the history-matching error defined by Equation (22). Figure 13 shows the history-matching results obtained by the machine-learning assisted workflow coupling ANN proxy models and PSO optimizer. It illustrates the fluid production matching quality predicted by the ANN models when the history-matching error was minimized. The average relative matching errors were 0.9%, 42.9%, and 17.2% for oil, water, and gas production rates comparing the real-field

historical data, respectively. Although the overall history-matching results obtained by the PSO algorithm indicated a 20.2% relative error, water production exhibited much worse performance compared to that of oil and gas. Thus, a confirmation by revisiting the high-fidelity numerical simulator was quite necessary.

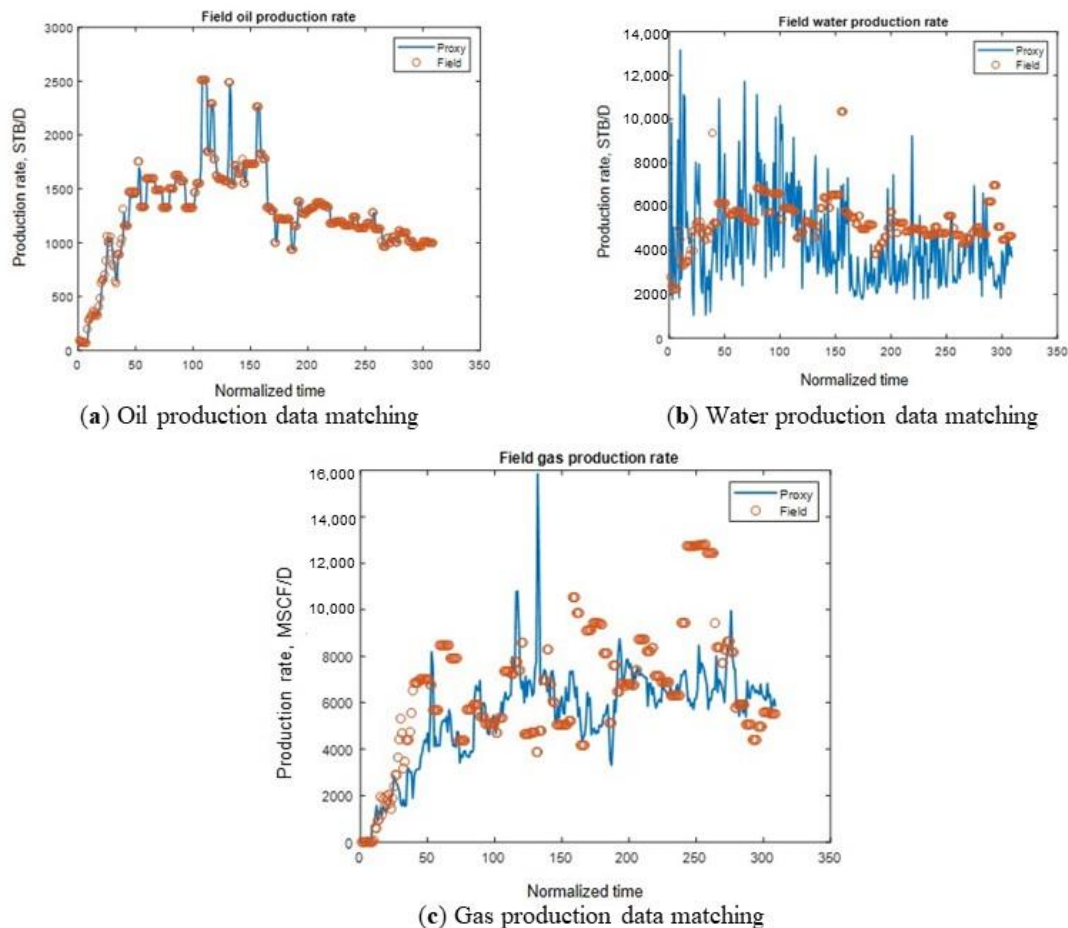


Figure 13. History-matching results obtained by the proxy model for (a) oil production, (b) water production, and (c) gas production data.

The major objective of the machine-learning assisted history-matching workflow was to find the combination of reservoir hydrodynamic properties that make the reservoir simulation model predictions agree with the field historical data. When the PSO optimizer minimizes the history-matching error during the iterative processes, the ANN proxies are employed to predict the field responses. Notably, a set of reservoir properties would be obtained after the PSO iteration converges, which is considered the solution of the history-matching process. The history matching solution must revisit the high-fidelity model for the following reasons:

1. Although the proxy models were well-trained, there existed potential error margins. A history-matching solution must feed into the high-fidelity simulator and confirm the matching quality.
2. To structure the forecasting scenarios, a base-case numerical simulation model needed to be established by re-running the high-fidelity simulator using the history matching solution suggested by the machine-learning assisted workflow.

Figure 14 shows the history-matching quality confirmed by the high-fidelity simulator considering the solution found by the machine-learning assisted workflow. Good agreements in terms of oil/gas/water production and the gas injection rate were observed, which indicates that the numerical model was well tuned and effectively characterized

the underground hydrodynamic environment. More importantly, the successful history-matching study stressed the robustness of the proxy model by learning the data structure presented by the dataset. The computational cost was significantly reduced by the machine-learning assisted workflow. Preparing the 100 simulation runs took 80 h of CPU time. With the help of the proxy model, the workflow completed more than 600 PSO iterations using a population size of 100 (60,000 realizations in total) within 300 s.

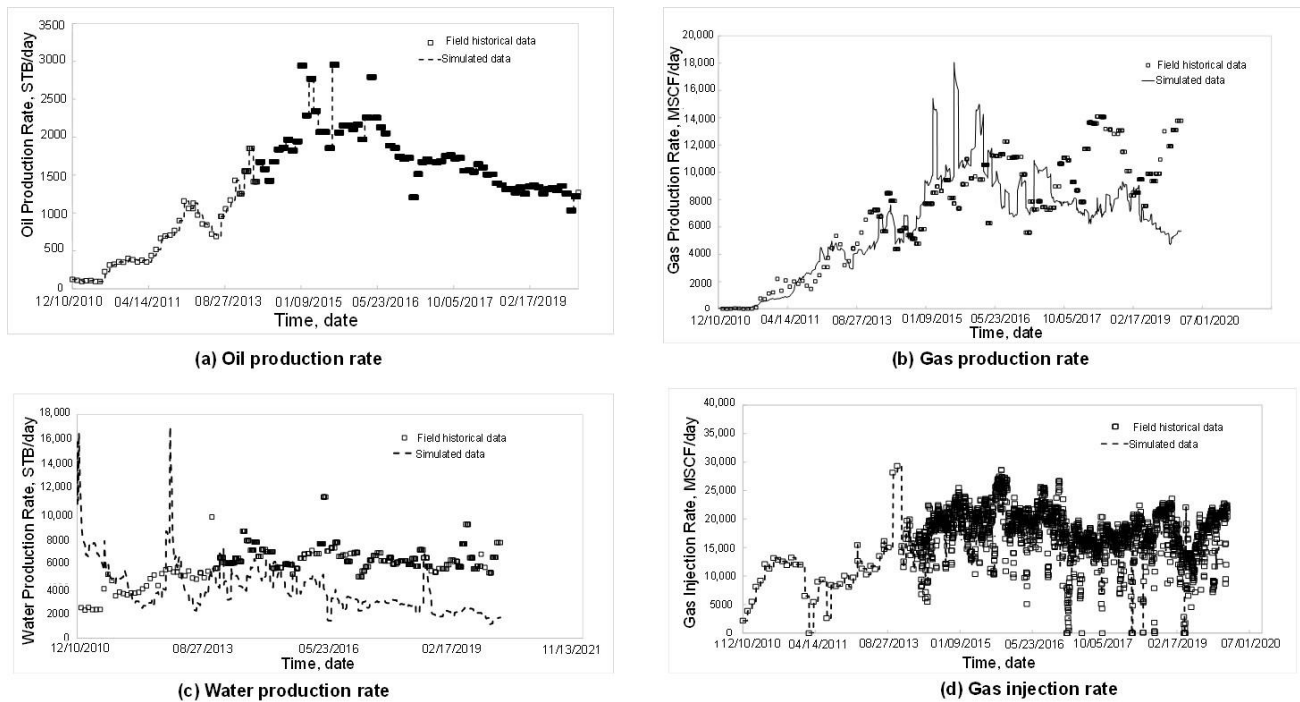


Figure 14. History-matching results confirmed using the high-fidelity numerical model for (a) the oil production rate, (b) the gas production rate, (c) the water production rate, and (d) the gas injection rate.

6.2. A Multi-Objective Optimization Application

Another application of the proposed workflow successfully designed the CO₂-WAG injection in the FWU field for the period from January 2020 to January 2038. The CO₂ injection includes the purchased CO₂ and the recycled CO₂ from the produced gas. According to the initial analysis of the simulation results and previous studies [31], as shown in Figure 15, the purchased CO₂ rate varies from 2020 to 2033. At the end of 2033, no more CO₂ is purchased.

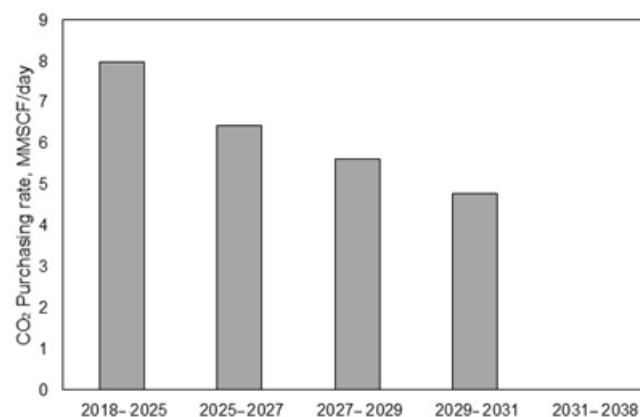


Figure 15. CO₂ purchasing rate.

A base-case forecasting scenario was established by dividing the CO₂-WAG injection patterns into four groups (WAG A, B, C, and D). The entire project timeline was split into eight time periods, as shown in Table 1. There are more wells added to the groups as the project processes.

Table 1. Time period details of the development plan.

Stages	Start	End
1	January 2020	June 2020
2	July 2020	January 2022
3	January 2022	January 2023
4	January 2023	January 2026
5	January 2026	January 2028
6	January 2028	January 2030
7	January 2030	January 2032
8	January 2032	January 2038

The CO₂-WAG injection parameters included the water and gas injection duration, water injection rates, and production well specifications, which vary for each group and time period. Thus, there were 37 design specifications considered in the optimization study. The NPV, oil recovery, and CO₂ storage volume were the objective functions. A base case model was structured using default design parameters suggested by the field operator [2]. The incremental oil production, NPV, and CO₂ injection/production volume of the base case model are displayed in Figure 16. The objective functions of the base case model are summarized in Table 2. Note that the CO₂ storage efficacy is the ratio of CO₂ sequestered to the total purchased volume.

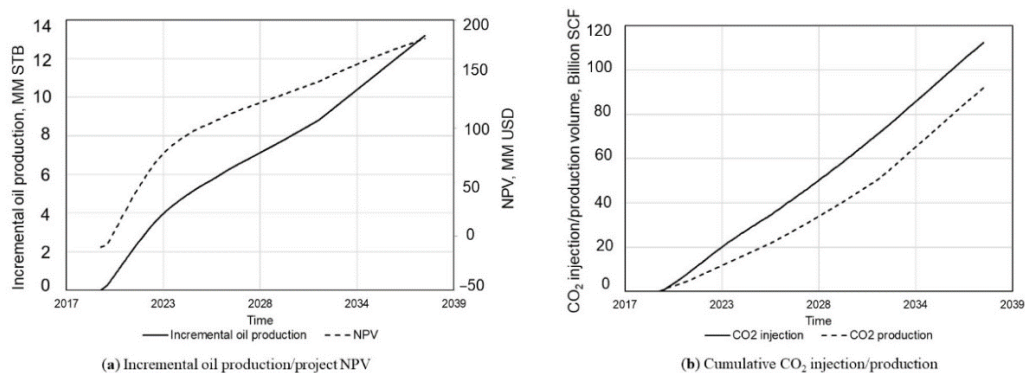


Figure 16. Simulation results of the base case model for (a) incremental oil production/project NPV and (b) cumulative CO₂ injection/production.

Table 2. Objective functions of the base case model.

Objective Function	Unit	Value
Oil production increments	MM bbl	13.3
CO ₂ storage volume increments	MM metric ton	1.06
Project NPV	Million USD	183
Cumulative Oil Production by January 2038	MM STB	16.8
Total CO ₂ storage volume by January 2038	MM metric ton	2.36
CO ₂ storage efficacy	percentage	81.19%

The cumulative oil recovery was 16.8 MM bbl and the total CO₂ storage was 2.36 million metric tons, which was 81.2% of the total purchased CO₂. The project NPV was USD 183 million. It is worth stressing that the average reservoir pressure of ≥ 4000 psi was

imposed as the physical constraint on the multi-objective optimization to sustain a miscible flooding process [6]. The proposed optimization workflow aims at improving all the objective functions simultaneously.

As shown in Figure 17, the Pareto front solution considering three objective functions is illustrated as a space curve in the domain of f_1 (oil recovery), f_2 (CO₂ storage), and f_3 (NPV). The dominating solution sitting on the Pareto front was validated by the numerical simulator. The good agreements can be observed in Figure 17, which confirms the validity of the proxy model and the MOPSO optimization results. Another interesting observation was drawn by plotting the projections of the three-objective Pareto front to three of the orthogonal surfaces. In Figure 18, a strong tradeoff relationship can be observed between the oil recovery and NPV, as well as the oil recovery and CO₂ storage volume, which means that by improving the oil recovery, the NPV and CO₂ storage volume has to be sacrificed. Therefore, the field operator could have a flexible range of choices to design the CO₂-WAG process based on the primary desire of the project. However, due to the tax credit brought by the CO₂ sequestration, the project NPV and CO₂ storage volume increases monotonically.

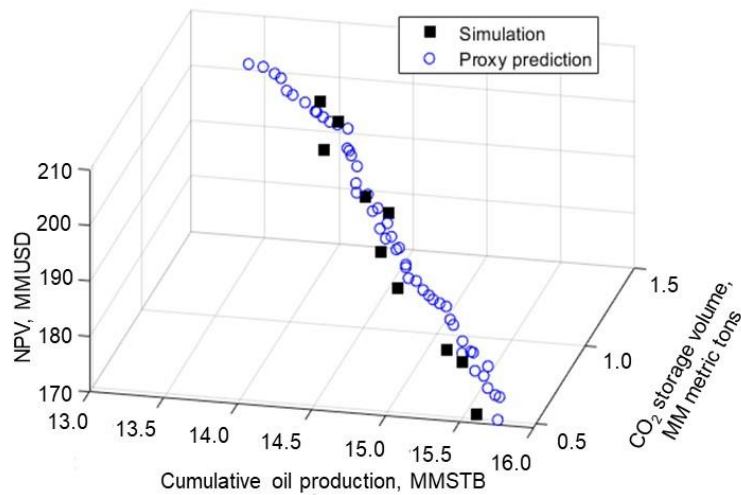


Figure 17. Comparison between the Pareto front and numerical simulation results using some of the optimized input parameters: three-objective.

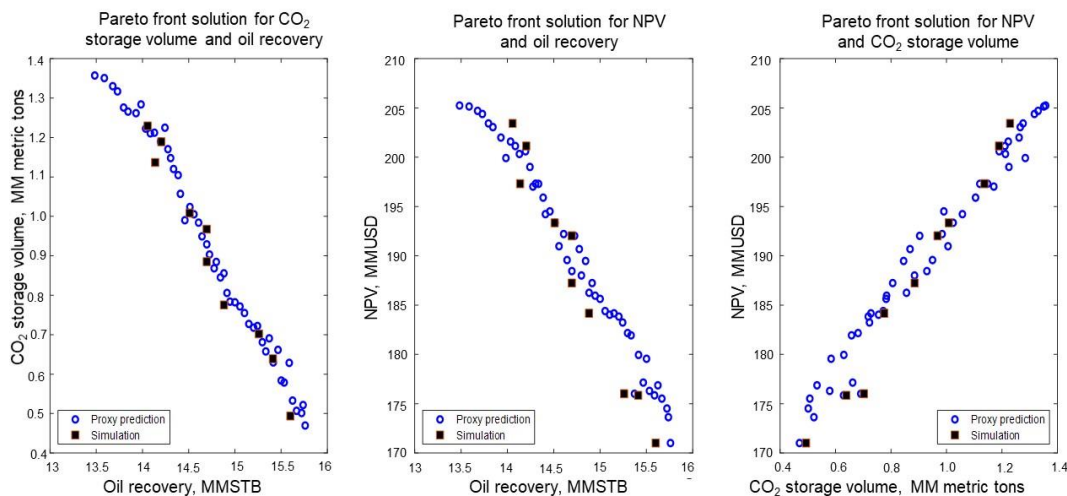


Figure 18. Projection view of the comparison between the Pareto front and numerical simulation results using some of the optimized input parameters: three-objective.

Table 3 summarizes the optimization results using the proposed workflow and the base case values. The base case oil recovery fell out of the range of the Pareto front solution, which means that the base case design scenario was part of the dominating solution. Notably, the range of oil recovery included in the Pareto front was quite narrow (13.3–15.7 MM STB), and the improvement to the base case was 16.0%. The design of CO₂-WAG made a considerable difference for the CO₂ storage volume and NPV. Therefore, the solution yielding the highest NPV and CO₂ storage volume, even with the lowest oil recovery, should be considered with priority. However, that investigation could be altered when the oil price increases, or the tax allowance brought by CO₂ utilization becomes lower.

Table 3. Comparison between the results of the base case and the optimized solutions.

Item	Unit	Range/Value	Base Case
Oil production increments	MM STB	13.3–15.7	13.2
CO ₂ storage volume increments	MM metric ton	0.42–1.4	1.06
Project NPV	Million USD	170–205	-
Max cumulative oil production	MM STB	19.3	16.8
Max cumulative CO ₂ storage	MM metric ton	2.7	3.63
Max project NPV	Million USD	205	183
CO ₂ storage efficacy	percentage	92.90%	81.19%

The field operator collaborating with the SWP project aims to extract hydrocarbon in the Morrow-B formation using CO₂-EOR technologies. The primary goal of the operation is to produce residual oil, and the sequestration of CO₂ during the EOR stage would help improve the project NPV via tax credits. That is why oil recovery, the NPV, and CO₂ storage volume are selected as the objective functions. The key to maximizing the CO₂ storage (minimizing the CO₂ production) during the EOR process is the injection specifications. The miscible flooding processes inject CO₂ gas in continuous or cyclic manners. For a multiphase flow system, the mobility of the gaseous phase relates to the gas saturation. As the gas injection processes, a high gas saturation channel could form from which the injected gas could break through. Such an issue would occur in the continuous CO₂ injection process. In the practical CO₂ injection process, the water and CO₂ may inject in a cyclic manner and avoid the persistent buildup of gas saturation. More importantly, since the pressure and saturation distributions of the system would exhibit strong heterogeneity, the CO₂-WAG design should vary for different injectors. For instance, injectors in the high-water saturation region could take a higher gas injection volume (longer gas injection cycle) and vice versa. Thus, an accurate characterization of the pressure and saturation distribution is also critical for a successful CO₂-WAG design. The deployment of CO₂ foam [36] exhibits more robust mobility control competences. Although these technologies are not used in the Morrow-B CO₂-EOR project, the coupling of chemical slug and CO₂ gas would be beneficial for CO₂ storage and oil recovery under feasible project economic criteria.

7. Conclusive Remarks

This paper presents a comprehensive summary of machine-learning assisted workflows to solve various engineering problems associated with the CO₂-WAG design. By coupling the proxy model and the advanced optimization protocols, the computational overheads of the history-matching and multi-objective optimization processes are significantly reduced. Besides the successful experiences obtained from this work, the following limitations cannot be ignored:

1. The proxy model developed in this work is a field-specified model that only works for the Morrow-B formation. Its implementation in other fields needs a new reservoir simulation model structure and needs to go through the proposed workflow.

2. The CO₂-EOR process only includes the CO₂-WAG process. Other CO₂-EOR technologies such as CO₂ foam, continuous injection, CO₂ huff-n-puff, etc., are not considered. From this work, the following conclusive statements can be drawn:
 1. The selection of the machine-learning algorithm may comprehensively consider the dimension of the problem and the demand of error margin. The RSM, SVM, and MLNN are suitable for different types of datasets and a wise choice of method could essentially enhance the prediction performance of the proxy model.
 2. Although machine-learning approaches exhibit many superiorities over the conventional numerical approach, a precise reservoir engineering analysis should take advantage of both. The error margin of the proxy model is the tradeoff for accelerating the computational speed. Thus, validation via the high-fidelity numerical model is necessary before deploying the results in operational practices.
 3. The Pareto front optimum protocol provides an alternative way to address multi-objective optimization problems. However, a successful application of the Pareto front optimum solution must be based on the tradeoff relationship between various objective functions.
 4. The calculation of the project economic objective functions strongly depends on the tax allowance and crude oil price. Therefore, the operational deployment of the optimum design suggested by the workflow needs to take practical considerations such as crude oil market condition, government policies, etc., into account.

SI-Field Unit Conversion Factor

ft ×	0.3048	= m
ft ² ×	0.0929030	= m ²
ft ³ ×	0.0283169	= m ³
bbl ×	0.1589873	= m ³
psi ×	6.894757	= kPa

Author Contributions: Conceptualization, Q.S. and W.A.; methodology, Q.S. and W.A.; software, Q.S. and J.Y.; validation, Q.S., W.A., and J.Y.; formal analysis, M.C. and R.B.; investigation, Q.S. and J.Y.; resources, W.A., M.C., and R.B.; data curation, W.A. and J.Y.; writing—original draft preparation, Q.S. and W.A.; writing—review and editing, Q.S., W.A., and M.C.; visualization, Q.S., J.Y., and W.A.; supervision, W.A.; project administration, M.C., W.A., and R.B.; funding acquisition, W.A. and R.B. All authors have read and agreed to the published version of the manuscript.

Funding: Funding for this project was provided by the U.S. Department of Energy’s (DOE) National Energy Technology Laboratory (NETL) through the Southwest Regional Partnership on Carbon Sequestration (SWP) under Award No. DE-FC26-05NT42591.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

ANN	artificial neural network
CAPEX	capital expenditure
CCUS	Carbon capture utilization and sequestration
CPU	central processing unit
EOR	enhanced oil recovery
FWU	Farnsworth Unit
GA	genetic algorithm

HFU	hydraulic flow unit
MLNN	multi-layer neural networks
MM	million
MOO	multi-objective optimization
MOPSO	Multi-objective particle swarm optimization
NPV	net present value
OGIP	original gas in place
OOIP	original oil in place
POP	population
PSO	particle swarm optimization algorithm
REP:	repository
res bbl	reservoir barrel
RSM	response surface models
SCF	standard cubic feet
STB	stock tank barrel
SVM	support vector machines
SWP	Southwest Regional Partnership on Carbon Sequestration
VEL	velocity

References

1. You, J.; Ampomah, W.; Kutsienyo, E.J.; Sun, Q.; Balch, R.S.; Aggrey, W.N.; Cather, M. Assessment of Enhanced Oil Recovery and CO₂ Storage Capacity Using Machine Learning and Optimization Framework. In Proceedings of the SPE Europec Featured at 81st EAGE Conference and Exhibition, London, UK, 3–6 June 2019; Society of Petroleum Engineers: Dallas, TX, USA, 2019.
2. Ampomah, W.; Balch, R.; Grigg, R.B.; Cather, M.; Gragg, E.; Will, R.A.; White, M.; Moodie, N.; Dai, Z. Performance assessment of CO₂-enhanced oil recovery and storage in the Morrow reservoir. *Géoméch. Geophys. Geo-Energy Geo-Resour.* **2017**, *3*, 245–263. [[CrossRef](#)]
3. Munson, T.W. Depositional, Diagenetic, and Production History of the Upper Morrowan Buckhaults Sandstone, Farnsworth Field, Ochiltree County Texas. *Shale Shak. Dig.* **1994**, XXXX–XXXXIV, 2–20.
4. Ampomah, W.; Balch, R.S.; Grigg, R.B.; Will, R.; Dai, Z.; White, M.D. Farnsworth field CO₂-EOR project: Performance Case History. In Proceedings of the SPE Improved Oil Recovery Conference, Tulsa, OK, USA, 11–13 April 2006; Society of Petroleum Engineers: Dallas, TX, USA, 2006.
5. Ampomah, W.; Balch, R.S.; Ross-Coss, D.; Hutton, A.; Cather, M.; Will, R.A. An integrated Approach for Characterizing a Sandstone Reservoir in the Anadarko Basin. In Proceedings of the Offshore Technology Conference, Houston, TX, USA, 2–5 May 2016.
6. Ampomah, W.; Balch, R.; Cather, M.; Will, R.; Gunda, D.; Dai, Z.; Soltanian, M. Optimum design of CO₂ storage and oil recovery under geological uncertainty. *Appl. Energy* **2017**, *195*, 80–92. [[CrossRef](#)]
7. Ampomah, W.; Balch, R.S.; Grigg, R.B.; McPherson, B.; Will, R.A.; Lee, S.-Y.; Dai, Z.; Pan, F. Co-optimization of CO₂-EOR and storage processes in mature oil reservoirs. *Greenh. Gases Sci. Technol.* **2017**, *7*, 128–142. [[CrossRef](#)]
8. Kutsienyo, E.J.; Ampomah, W.; Sun, Q.; Balch, R.S.; You, J.; Aggrey, W.N.; Cather, M. Evaluation of CO₂-EOR Performance and Storage Mechanisms in an Active Partially Depleted Oil Reservoir. In Proceedings of the SPE Europec featured at 81st EAGE Conference and Exhibition, London, UK, 3–6 June 2019; Society of Petroleum Engineers: Dallas, TX, USA, 2019.
9. Ampomah, W.; Balch, R.S.; Grigg, R.B. Analysis of upscaling algorithms in heterogeneous reservoirs with different recovery processes. In Proceedings of the SPE Production and Operations Symposium, Oklahoma City, OK, USA, 1–5 March 2015; Society of Petroleum Engineers: Dallas, TX, USA, 2015.
10. Dai, Z.; Viswanathan, H.; Middleton, R.; Pan, F.; Ampomah, W.; Yang, C.; Jia, W.; Xiao, T.; Lee, S.-Y.; McPherson, B. CO₂ Accounting and Risk Analysis for CO₂ Sequestration at Enhanced Oil Recovery Sites. *Environ. Sci. Technol.* **2016**, *50*, 7546–7554. [[CrossRef](#)] [[PubMed](#)]
11. Pan, F.; McPherson, B.J.; Dai, Z.; Jia, W.; Lee, S.-Y.; Ampomah, W.; Viswanathan, H.; Esser, R. Uncertainty analysis of carbon sequestration in an active CO₂-EOR field. *Int. J. Greenh. Gas Control* **2016**, *51*, 18–28. [[CrossRef](#)]
12. Ahmmed, B.; Appold, M.S.; Fan, T.; McPherson, B.J.O.L.; Grigg, R.B.; White, M.D. Chemical effects of carbon dioxide sequestration in the Upper Morrow Sandstone in the Farnsworth, Texas, hydrocarbon unit. *Environ. Geosci.* **2016**, *23*, 81–93. [[CrossRef](#)]
13. Ross-Coss, D.; Ampomah, W.; Cather, M.; Balch, R.S.; Mozley, P.; Rasmussen, L. An Improved Approach for Sandstone Reservoir Characterization. In Proceedings of the SPE Western Regional Meeting, Anchorage, AK, USA, 23–26 May 2016; Society of Petroleum Engineers: Dallas, TX, USA, 2016.
14. You, J.; Ampomah, W.; Sun, Q.; Kutsienyo, E.J.; Balch, R.S.; Dai, Z.; Cather, M.; Zhang, X. Machine learning based co-optimization of carbon dioxide sequestration and oil recovery in CO₂-EOR project. *J. Clean. Prod.* **2020**, *260*, 120866. [[CrossRef](#)]
15. Ketineni, S.P.; Ertekin, T.; Anbarci, K.; Sneed, T. Structuring an Integrative Approach for Field Development Planning Using Artificial Intelligence and its Application to an Offshore Oilfield. In Proceedings of the SPE Annual Technical Conference and Exhibition, Houston, TX, USA, 28–30 September 2015; Society of Petroleum Engineers: Dallas, TX, USA, 2015.

16. Zhang, Z.; Jung, H.Y.; Datta-Gupta, A.; Delshad, M. History Matching and Optimal Design of Chemically Enhanced Oil Recovery Using Multi-Objective Optimization. In Proceedings of the SPE Reservoir Simulation Conference, Galveston, TX, USA, 10–11 April 2019; Society of Petroleum Engineers: Dallas, TX, USA, 2019.
17. Zhang, X.; Wu, J.; Coutier-Delgosha, O.; Xiao, H. Recent progress in augmenting turbulence models with physics-informed machine learning. *J. Hydrodyn.* **2019**, *31*, 1153–1158. [[CrossRef](#)]
18. You, J.; Ampomah, W.; Sun, Q. Development and application of a machine learning based multi-objective optimization workflow for CO₂-EOR projects. *Fuel* **2020**, *264*, 116758. [[CrossRef](#)]
19. Ertekin, T.; Ayala, L.F. *Reservoir Engineering Models: Analytical and Numerical Approaches*, 1st ed.; McGraw-Hill Education: Columbus, OH, USA, 2018; pp. 1–368.
20. Putcha, V.B.; Ertekin, T.A. Fast and Robust Compositional, Multi-Phase, Non-Isothermal Wellbore Hydraulics Model for Vertical Wells. In Proceedings of the SPE Annual Technical Conference and Exhibition, San Antonio, TX, USA, 9–11 October 2017; Society of Petroleum Engineers: Dallas, TX, USA, 2017.
21. Ayala, L.F.; Ertekin, T.; Adewumi, M.A. Study of Gas/Condensate Reservoir Exploitation Using Neurosimulation. *SPE Reserv. Eval. Eng.* **2007**, *10*, 140–149.
22. Panja, P.; Velasco, R.; Pathak, M.; Deo, M. Application of artificial intelligence to forecast hydrocarbon production from shales. *Petroleum* **2018**, *4*, 75–89. [[CrossRef](#)]
23. Sumardi, H.R.; Irawan, D. Coalbed Methane Production Parameter Prediction and Uncertainty Analysis of Coalbed Methane Reservoir with Artificial Neural Networks. In Proceedings of the Indonesian Petroleum Association Fortieth Annual Convention and Exhibition, Jakarta, Indonesia, 25–27 May 2016; American Association of Petroleum Geologists: Tulsa, OK, USA, 2016.
24. Parada, C.H.; Ertekin, T.A. New Screening Tool for Improved Oil Recovery Methods Using Artificial Neural Networks. In Proceedings of the SPE Western Regional Meeting, Bakersfield, CA, USA, 21–23 March 2012; Society of Petroleum Engineers: Dallas, TX, USA, 2012.
25. Thararoop, P.; Karpyn, Z.; Gitman, A.; Ertekin, T. Integration of seismic attributes and production data for infill drilling strategies—A virtual intelligence approach. *J. Pet. Sci. Eng.* **2008**, *63*, 43–52.
26. Li, Y.; Zhang, T.; Sun, S. Acceleration of the NVT Flash Calculation for Multicomponent Mixtures Using Deep Neural Network Models. *Ind. Eng. Chem. Res.* **2019**, *58*, 12312–12322. [[CrossRef](#)]
27. Rasmussen, L.; Fan, T.; Rinehart, A.; Luhmann, A.; Ampomah, W.; Dewers, T.; Heath, J.; Cather, M.; Grigg, R. Carbon Storage and Enhanced Oil Recovery in Pennsylvanian Morrow Formation Clastic Reservoirs: Controls on Oil–Brine and Oil–CO₂ Relative Permeability from Diagenetic Heterogeneity and Evolving Wettability. *Energies* **2019**, *12*, 3663. [[CrossRef](#)]
28. Bishop, C.M. *Pattern Recognition and Machine Learning*, 1st ed.; Springer: New York, NY, USA, 2006; pp. 1–738.
29. Hassoun, M. *Fundamentals of Artificial Neural Networks*; MIT Press: Cambridge, MA, USA, 2003; pp. 1–540.
30. Smola, A.J.; Schölkopf, B. A tutorial on support vector regression. *Stat. Comput.* **2004**, *14*, 199–222. [[CrossRef](#)]
31. You, J. Multi-Objective Optimization of Carbon Dioxide Enhanced Oil Recovery Projects Using Machine Learning Algorithms. Ph.D. Thesis, New Mexico Institute of Mining and Technology, Socorro, NM, USA, 2020.
32. Coello, C.; Pulido, G.; Lechuga, M. Handling multiple objectives with particle swarm optimization. *IEEE Trans. Evol. Comput.* **2004**, *8*, 256–279. [[CrossRef](#)]
33. Kennedy, J.; Eberhart, R. Particle swarm optimization. In Proceedings of the ICNN'95—International Conference on Neural Networks, Perth, WA, Australia, 27 November–1 December 1995.
34. Sun, Q. The Development of an Artificial-Neural-Network-Based Toolbox for Screening and Optimization of Enhanced oil Recovery Projects. Ph.D. Thesis, University Park, The Pennsylvania State University, State College, PA, USA, 2017.
35. Sun, Q.; Ertekin, T. The Development of Artificial-neural-network-based Universal Proxies to Study Steam Assisted Gravity Drainage (SAGD) and Cyclic Steam Stimulation (CSS) Processes. In Proceedings of the SPE Western Regional Meeting, Garden Grove, CA, USA, 27–30 April 2015; Society of Petroleum Engineers: Dallas, TX, USA, 2015.
36. Ibrahim, A.; Emrani, A.; Nasraddin, H. Stabilized CO₂ Foam for EOR Applications. In Proceedings of the Carbon Management Technology Conference, Houston, TX, USA, 17–20 July 2017.