Journal of Petroleum Science and Engineering 220 (2023) 111166

Contents lists available at ScienceDirect

Journal of Petroleum Science and Engineering

journal homepage: www.elsevier.com/locate/petrol

Optimizing initiation time of waterflooding under geological uncertainties with Value of Information: Application of simulation-regression approach

Cuthbert Shang Wui Ng *, Ashkan Jahanbani Ghahfarokhi

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

ARTICLE INFO

Keywords:
Decision analysis
Waterflooding
Value of information
Simulation-regression approach
Machine learning
Optimization under uncertainty

ABSTRACT

Reservoir Management (RM) is an example of sequential decision problems in the oil and gas industry. Therefore, implementing Decision Analysis (DA) tool to systematically resolve such problems has been a common practice. The value of Information (VOI) framework acts as one of these tools that helps reservoir engineers to manage RM problems. Regarding this, the Least-Squares Monte Carlo (LSM) algorithm, which is one of the simulation-regression approaches, has been employed to estimate VOI for a better quality of decision-making (DM). Integration of the LSM algorithm in RM is coined as "Sequential Reservoir Decision-Making" (SRDM). This approximate method is essential to resolve a sequential decision problem with high dimensionality caused by many possible outcomes of uncertainties. This challenge is generally known as the "curse of dimensionality". In this work, a modified LSM algorithm has been applied under the SRDM paradigm to optimize the waterflooding initiation time considering geological uncertainties. The modification considers the effects of information acquired previously and at the current decision time before a decision is made. The reservoir model used in this work is the OLYMPUS benchmark model. Apart from utilizing Linear Regression (LR) in the LSM algorithm, the use of two machine learning (ML) techniques, viz. Gaussian Process Regression (GPR) and Support Vector Regression (SVR), have been illustrated to estimate the VOI. Based on the results, LR, GPR, and SVR correspondingly estimate the VOI as 11.52 million USD, 11.17 million USD, and 12.46 million USD. This means that SVR displays an improvement of 8.18% compared to the VOI assessed by LR. This shows its good applicability in VOI estimation and it can be concluded that integrating ML techniques into the SRDM paradigm demonstrates high potential for RM applications.

1. Introduction

Decision Analysis (DA) is one of the knowledge domains that has been ubiquitous in different aspects of engineering studies. According to Howard (1980), DA can be understood as a systematic methodology that transforms an opaque (hard to understand) decision problem into a transparent (easy to perceive) one via a series of transparent steps. Concerning this, Value of Information (VOI) is one of the most prevalent decision-making (DM) tools. VOI is the approximation of additional value induced when information is brought to a decision problem (Howard, 1966). Despite having such a lucid definition of DA, many engineers are still subject to misconception. They tend to include as many details as possible when they are developing their DM tool, including VOI. This might not be a good practice because only important or pertinent factors should be considered in DM models.

Furthermore, it is enlightening for engineers to realize that the VOI technique is formulated to evaluate if the improvement in DM by acquiring the information is worth the cost required to gain it. In another word, the VOI analysis is an a priori analysis that quantitatively assesses the benefits of obtaining additional information before the data is gathered and a decision is made (Hong et al., 2018). As Bratvold and Begg (2010) have counseled, for an information-gathering activity to be worthwhile, its VOI should exceed the cost of the activity itself. Also, it must have the ability to change the decision maker's beliefs about uncertainty and the decisions made otherwise. Hence, engineers ought to be cognizant that VOI does not in fact "reduce uncertainty", but it facilitates the adjustment of the decisions concerning underlying uncertainty. Thus, VOI is often coupled with uncertainty and DM, in which information cannot be valued without a specific decision context (Bratvold et al., 2009; Hong et al., 2018).

The use of the VOI methodology has been growing in the oil and gas industry, especially in the aspects of reservoir management (RM), for the

* Corresponding author.
E-mail address: cuthbert.s.w.ng@ntnu.no (C.S.W. Ng).

https://doi.org/10.1016/j.petrol.2022.111166
Received 21 April 2022; Received in revised form 7 October 2022; Accepted 24 October 2022
Available online 28 October 2022
0920-4105/

past decade. RM refers to the employment of available technology, labor resources, and financial assets to maximize economic returns through hydrocarbon production from a reservoir (Satter et al., 1998; Wiggins and Startzman, 1998). RM generally entails a series of operations and decisions, stemming from the initial phase of field discovery to the final phase of field abandonment. Furthermore, most of the DM problems are considered sequential and involve a lot of uncertainties. This implies that information is continuously acquired to enhance the quality of DM. Therefore, the VOI framework aptly applies in the resolution of such sequential DM problems. Nevertheless, the real-world challenge in this context is to determine the analytical solution of VOI.

One of the methods to approximately compute the VOI is by applying a decision tree (DT), which is a part of dynamic programming. DT is efficient for visualization and communication of the frame of a sequential decision. Fundamentally, a sequential decision problem can be represented as a DT and solved by rolling back the DT itself. For a more comprehensive implementation of DT, refer to these books (Bratvold and Begg, 2010; Howard and Abbas, 2016). Unfortunately, the DT method will encounter the “curse of dimensionality” if it is used to solve a more sophisticated decision problem (Powell, 2011). In this aspect, three main sources of the “curse of dimensionality” comprise the number of possible outcomes (or uncertainties), the number of decision points (time where a decision needs to be made), and the number of alternatives at each decision point (Powell, 2011). Least-Squares Monte Carlo (LSM) algorithm that was developed by Longstaff and Schwartz (2001) can replace DT in resolving a more complex problem, but it is only efficient to handle sequential decision problems with many uncertain quantities and limited number of alternatives. The increase in the number of alternatives and decision points causes an exponential increase in computational effort and thus, the “curse of dimensionality” arises.

LSM is placed under the umbrella of the simulation-regression approach in terms of the determination of VOI. As the name of the approach implies, it can be perceived that there are two main frameworks, namely simulation and regression analysis. Monte Carlo simulation (MCS) is one of the standard practices to capture the effect of uncertainties on the production profile. In reservoir engineering, uncertainties generally pertain to the geological properties of a reservoir. Hence, numerical reservoir simulation (NRS) can leverage MCS to perform forward modeling to generate production data under uncertainties for any RM decision problem. Thereafter, regression analysis is conducted in the form of backward calculation to estimate the VOI. The details of this analysis will follow later. Linear Regression (LR) has been used to perform the regression analysis. However, as the research domain has been developing, modifications or improvements to the LSM algorithm[1] have been done to resolve different sequential decision problems. More detailed descriptions will follow.

The application of VOI analysis in the oil and gas industry has been overviewed by Bratvold et al. (2009). In addition, several articles have illustrated the implementation of the simulation-regression approach under the VOI paradigm. Willigers and Bratvold (2009) performed the valuation of real options in an oil and gas project through the implementation of LSM. Stemming from this work, LSM was further employed for the valuation of swing contracts in the field of natural gas and electricity (Willigers et al., 2011). Alkhatib et al. (2013) discussed the use of LSM to yield an optimal policy of surfactant flooding in both homogeneous and heterogeneous reservoirs considering geological uncertainty. Hong et al. (2019) further extended the use of LSM by coupling this algorithm with a proxy model, known as the Two-Factor Production Model that was comprehensively discussed (Parra Sanchez, 2010) to evaluate the optimal switch time of waterflooding. The LSM algorithm was modified to integrate the dependency of both currently and previously measured data. Therefore, the algorithm was named modified LSM. Based upon the formulation of modified LSM, Tadjer et al. (2021a) evaluated the VOI under polymer flooding. Besides LR, they successfully utilize machine learning (ML) techniques (nonlinear regression), including neural network regressor and Tree-based Pipeline Optimization Tool (TPOT), which was proposed by Olson et al. (2016), as an alternative. Dutta et al. (2019a) also displayed how Principal Component Regression and Partial Least-Squares Regression could be implemented as a nonlinear regression approach to assess VOI for sequential spatial data collection in subsurface energy application. There are also other papers (Dutta et al., 2019b; Eidsvik et al., 2017) expounding on the application of simulation-regression approaches for the estimation of VOI. Furthermore, these nonlinear simulation-regression approaches have been illustrated in other interesting tasks with the emphasized application in Carbon, Capture, and Storage (CCS). Tadjer et al. (2021b) implemented TPOT as the regression technique to determine the VOI of performing carbon storage in Utsira formation. Also, Anyosa et al. (2021) applied some ML-based regression methods, including k-Nearest Neighbors, Random Forest, and Convolution Neural Networks, to do VOI analysis to evaluate the value of seismic monitoring of CO_2 storage at Smeaheia site.

The work that is conducted here is inspired by a previous work (Ng, 2019). In this paper, the modified LSM algorithm is implemented to determine the optimal initiation time of waterflooding in the OLYMPUS reservoir model under geological uncertainties. This initiation is decided based on the acquisition of information from both oil and water production data. Moreover, 50 different geological realizations have been employed to capture the uncertainties in the DM process. The pertinent details will follow in later sections. Apart from the conventional LR approach for regression analysis, other nonlinear regression techniques are also utilized. Examples of the nonlinear techniques (alternatively termed ML-based methods) chosen in this work consist of Gaussian Process Regression and Support Vector Regression. The corresponding computed VOI and the decisions to be made for each geological realization by incorporating different regression methods are then analyzed and compared for further discussion.

After this introduction, the paper is structured by having the following sections. Section 2 provides the theoretical framework of VOI in which the mathematical implementation of VOI estimation is presented. The background of the decision problem and the details of the OLYMPUS reservoir model as well as the economic model employed are thereafter briefed under Section 3. Then, Section 4 discusses the mechanism of the modified LSM along with its integration into “Sequential Reservoir Decision-Making” (SRDM). Section 5 explains the other nonlinear regression ML-based methods used in this work. Thereafter, Section 6 highlights the results and relevant discussions. Some concluding remarks are summarized in Section 7.

2. Value of Information (VOI)

In any information acquisition activity, VOI relies upon two important uncertainties, namely distinction of interest and observable distinction (Bratvold et al., 2009). The distinction of interest is not observable and aimed to be learned. Therefore, any information obtained in the form of any test result is considered as the observable distinction that helps the decision makers to perceive better the distinction of interest. In the context of RM, specifically production optimization, the production data gained until time t (when the decision is to be made) is treated as an observable distinction. It is computationally challenging to analytically represent the distribution of observable distinction due to its high dimension. Therefore, the use of

[1] LSM is precisely a combined application of MCS and LR. The use of sampling techniques other than MCS for the generation of different realizations of simulation and other data-driven methods as substitutes for LR is better termed as simulation-regression method.

Monte Carlo sampling plays a role to remediate this issue. Based on the assumption of risk neutrality[2], VOI can be mathematically represented as follows:

$$\text{Value of Information} = \max\{0, \gamma\}$$
$$\gamma = \left[\begin{array}{c}\text{Expected Value with}\\ \text{Information}\end{array}\right] - \left[\begin{array}{c}\text{Expected Value without}\\ \text{Information}\end{array}\right] \tag{1}$$

The estimated γ under a decision problem can be negative. Negative γ denotes that it is not economically feasible to acquire information. Hence, the lower limit of VOI is always treated as zero. Besides that, for Expected Value without Information (EVWOI), the corresponding decision without information (DWOI) is the alternative that optimizes the EV over all the realizations. For Expected Value with Information (EVWI), the respective optimal decision is Decision with Information (DWI). The mathematical formulations of EVWOI and EVWI are respectively displayed as:

$$\text{EVWOI} = \max_{\mathbf{a}\in A}\left[\int \mu(\mathbf{x},\mathbf{a})\, p(\mathbf{x})\, d\mathbf{x}\right] \approx \max_{\mathbf{a}\in A}\left(\frac{1}{N_r}\sum_{r=1}^{N_r}\mu(\mathbf{x}^r,\mathbf{a})\right)$$
$$\text{where}\quad \mathbf{a}_{\mathbf{DWOI}}^{\mathbf{optimal}} = \arg\max_{\mathbf{a}\in A}\left(\frac{1}{N_r}\sum_{r=1}^{N_r}\mu(\mathbf{x}^r,\mathbf{a})\right) \tag{2}$$

$$\text{EVWI} = \int \max_{\mathbf{a}\in A}\left[E(\mu(\mathbf{x},\mathbf{a})|\mathbf{y})\right] p(\mathbf{y})\, d\mathbf{y} \approx \frac{1}{N_r}\sum_{r=1}^{N_r}\max_{\mathbf{a}\in A} E[(\mu(\mathbf{x},\mathbf{a})|\mathbf{y}^r)]$$
$$\text{where}\quad \mathbf{a}_{\mathbf{DWI}}^{\mathbf{optimal}} = \arg\frac{1}{N_r}\sum_{r=1}^{N_r}\max_{\mathbf{a}\in A} E[(\mu(\mathbf{x},\mathbf{a})|\mathbf{y}^r)] \tag{3}$$

According to the formulations above, $p(\mathbf{x})$ is a prior probability distribution of distinction of interest that is represented as an ensemble of $\mathbf{x} = \{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^{N_r}\}$. $\mathbf{a}$ is used to denote the available alternatives, which are from a set of possible alternatives, A. Furthermore, $\mu(\mathbf{x}^r, \mathbf{a})$ is the function that yields the prospect values corresponding to a specific realization and selected alternatives. $\mathbf{y}$ is a collection of observable data, in which $\mathbf{y} = \{\mathbf{y}^1, \mathbf{y}^2, \ldots, \mathbf{y}^{N_r}\}$ and $p(\mathbf{y})$ is the marginal probability distribution. For each realization of $\mathbf{x}^r$, forward modeling can be done to determine $\mathbf{y}^r$.

VOI can also be understood as VOII (Value of Imperfect Information) because it is very challenging to acquire perfect information regarding a DM context in real life. Information is perfect if it is always true. Equation (3) portrays the estimation of EVWII. Thus, in RM, perfect information is the information that reveals the true properties of a reservoir and the impacts of the recovery mechanism. Besides that, the value of perfect information (VOPI), which is the difference between Expected Value with Perfect Information (EVWPI) and EVWOI, acts as the upper limit of VOI. In this context, the decision with perfect information (DWPI) corresponds to an alternative that optimizes the relevant objective function for each realization based upon prior distributions. Finding the average of such values of the objective function over all the realizations yields EVWPI as follows:

$$\text{EVWPI} = \int \max_{\mathbf{a}\in A}[\mu(\mathbf{x},\mathbf{a})]\, p(\mathbf{x})\, d\mathbf{x} \approx \frac{1}{N_r}\sum_{r=1}^{N_r}\max_{\mathbf{a}\in A}[\mu(\mathbf{x}^r,\mathbf{a})]$$
$$\text{where}\quad \mathbf{a}_{\mathbf{DWPI}}^{\mathbf{optimal}} = \arg\frac{1}{N_r}\sum_{r=1}^{N_r}\max_{\mathbf{a}\in A}\mu(\mathbf{x}^r,\mathbf{a}) \tag{4}$$

[2] Risk neutrality is a risk attitude apart from risk-averse and risk-seeking. Please refer to this literature (Hillson and Murray-Webster, 2017) for more explanation of risk attitudes. In a simpler term, a risk-neutral decision maker applies the Expected Value (EV) in the process of DM. This implies that the decision maker will have the same preference over two alternatives with the same EV.

3. Background of the decision problem and models

3.1. Problem setting

The decision problem discussed here is a part of RM and similar problems have been briefed in several pieces of literature (Hong et al., 2019; Ng, 2019; Tadjer et al., 2021a). Fundamentally, this decision problem involves the optimization of the initiation time of waterflooding in a 3D reservoir model (the benchmark model OLYMPUS). In the framework of this sequential decision problem, the production period of the OLYMPUS model is assumed to be 10 years. Thereafter, each year, a decision is needed if it is better to switch from primary recovery to waterflooding (in other term, to start waterflooding) or continue only with primary recovery. The termination time of production (under both primary recovery and waterflooding) is then optimized too. Concerning these, the initiation of waterflooding and termination of production can only occur once.

3.2. Reservoir model

The reservoir model implemented in this paper is the OLYMPUS model and simulation is performed by using the Eclipse 100 software (Schlumberger, 2019). This benchmark case, a synthetic field model developed by Fonseca et al. (2020), mainly consists of an oil-water system and has an approximate dimension of 9 × 3 km. The geological properties of this model have a typical resemblance to those of the North Sea field with Brent-type oil. The model has a thickness of 50 m with two different zones separated by an impermeable shale layer. In addition, the model is made up of 341,728 grid blocks in which the average dimension of each block is 50 × 50 × 3 m. However, the total number of active grid blocks are 192,750.

Moreover, to resolve the sequential decision problem as explained earlier, an ensemble of 50 realizations is used to capture the effect of uncertainty in this context. The uncertain variables consist of facies, porosity, permeability, net-to-gross ratio, initial water saturation, and transmissibility across the faults. For further details of the geological and petrophysical aspects of OLYMPUS, please peruse Fonseca et al. (2020). About the well configuration in this model, there are 7 injectors and 11 producers. Each of the injectors is controlled by keeping the maximum well rate of 2000 sm^3/day with bottomhole pressure target of 250 bars. Besides that, each of the producers is controlled by having the maximum bottomhole pressure at 150 bars. With these sets of control, the initiation time of all the injectors is optimized by applying the VOI analysis. The architecture (permeability in x-direction, PERMX in the unit of mD) of one of the realizations of the OLYMPUS model used here is presented in Fig. 1.

3.3. Economic model

The economic model used in this work is represented by net present value (NPV), which is illustrated as follows:

$$\text{NPV} = \sum_{i=0}^{N_t}\frac{\Delta t_i\left(P_o q_o^i - P_w q_w^i - P_{wi} q_{wi}^i\right) - \text{CAPEX}_i}{(1+\text{interest rate})^{t_i}} \tag{5}$$

Based on the NPV equation above, P indicates the price in which the subscripts o, w, and wi respectively mean oil, water produced, and water injected. q^i indicates the production (or injection) rates at timestep i. Δt_i is the difference between timesteps i and i-1. The timestep is on yearly basis. CAPEX denotes capital expenditure. In addition, the values of economic variables applied in this paper are tabulated in Table 1. Based on Table 1, it can be noted that in the case of waterflooding, there are three types of CAPEX, such as capital expenditure for having only primary recovery, additional capital expenditure for starting waterflooding after primary recovery, and capital expenditure for starting waterflooding without having primary recovery.

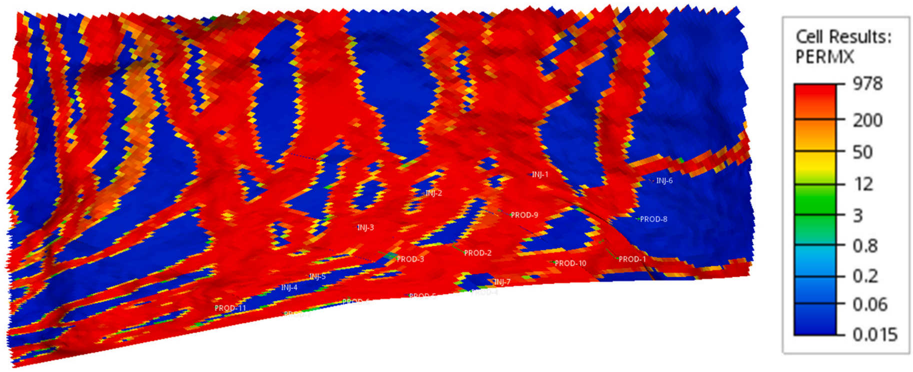

Fig. 1. Top view of Realization 5 of OLYMPUS.

Table 1
Values of economic variables used in this work.

Variables	Values	Units
Oil Price	408.85	USD/m^3
Water Production Price	50.32	
Water Injection Price	50.32	
CAPEX (Primary Recovery)	40	million USD
Additional CAPEX	30	
CAPEX (Waterflooding)	85	
Interest Rate	6%	per year

4. Simulation-regression paradigm

In general, EVWII under a specific decision context can be estimated by using the simulation-regression approach. This approach leverages MCS and regression analysis to determine the conditional EV upon given data. By implementing MCS, it is possible to alleviate the curse of dimensionality induced by the number of uncertain outcomes. Upon the completion of MCS, the backward induction is conducted with the aid of regression analysis to approximate the conditional EV of each alternative. The following are the details of the mechanisms corresponding to MCS and regression analysis (backward induction):

Monte Carlo simulation (MCS):

1) MCS is implemented to generate several realizations with different state variables (for instance, permeability, porosity, net-to-gross ratio, and so on) that can be understood as $\mathbf{x}^\mathbf{r}$.
2) Forward modeling of these realizations is done to create data (corresponding to oil/water production rates as well as water injection rates) to which noise will be added by using the statistics of the measurement errors. The noise is modeled by using zero mean and standard deviation of 0.15 here.
3) NPV of each decision alternative $\mu(\mathbf{x}^\mathrm{r}, \mathbf{a})$ is computed and EVWOI is thereafter determined by using equation (2).

Regression Analysis:

1) Beginning from the last decision point in time, NPVs are regressed on the generated data $\mathbf{y}^\mathbf{r}$ (considering only q_o and q_w) to determine the expected NPV (ENPV) of decision alternative **a** being conditional on the data, $\mathbf{y}^\mathbf{r}$. This procedure corresponds to the calculation of the term $\mathrm{E}[(\mu(\mathbf{x}, \mathbf{a})|\mathbf{y}^\mathrm{r})]$ as demonstrated on the right-hand side of equation (3).
2) The step above is repeated for every decision alternative.
3) The best decision alternative, $\mathbf{a}_{\mathbf{DWII}}^{\mathbf{optimal}}$ is made by selecting the decision alternative that yields the maximum conditional ENPV given the known data for every realization, and EVWII is computed based on equation (3). VOI is then calculated by using equation (1).

4.1. Sequential Reservoir Decision-Making (SRDM)

Regarding the details of regression techniques used, if the technique employed is the least-squares method (or LR), then it is termed the LSM algorithm as mentioned before. In this aspect, Hong et al. (2019) have discussed the application of LSM in the resolution of the initiation time of Improved Oil Recovery as an epitome of SRDM. On closer scrutiny, as LSM is employed for SRDM, the termination time given a specific initiation time of waterflooding must be first determined. Due to its nature of backward induction, the algorithm commences in Year 10.

In (the beginning of) Year 10, the optimal termination time is found by assuming that waterflooding has been started at this time. Therefore, there are only two available options (or decision alternatives), which are "terminate in Year 10" and "continue with waterflooding in Year 10". Therefore, the NPVs corresponding to these two options are regressed on the production data ranging from Year 1 to Year 9 given that waterflooding has started in Year 10. Thereafter, between these two options, that of higher estimated NPV is the optimal option in this case for each realization. Averaging the NPVs of these options over all realizations results in the ENPV of waterflooding initiation in (the beginning of) Year 10.

When the time rolls back to (the beginning of) Year 9, given waterflooding started the same year, there are three available options, namely "terminate in Year 9", "continue with waterflooding in Year 9 but terminate in Year 10", and "continue with waterflooding at Year 10". The last option corresponds to those determined in the previous step. Hence, the NPVs of the last two options are first regressed on the production data from Year 1 to Year 9 (these two options are regressed first due to the availability of data from Year 1 to Year 9). Based upon these estimated NPVs, the two options are compared and the respective optimal NPV is recorded for each realization. Then, the chosen option for every realization is compared with the option of "terminate in Year 9" through another regression analysis using the production data ranging from Year 1 to Year 8. This whole step will determine the ENPV of waterflooding initiation in (the beginning of) Year 9. The same logic is conducted every previous year. This procedure assists us to select the optimal stopping time for each year by assuming that waterflooding is initiated in that particular year.

Upon completing this procedure, the optimal option of waterflooding initiation considering termination has been determined Then, these options are compared with the option of "continuing only with primary recovery". In this case, in (the beginning of) Year 10, the NPVs of "initiating waterflooding in Year 10 with its respective optimal termination time" and "continuing only with primary recovery" are regressed on the production data from Year 1 to Year 9. The higher approximated NPV is then used to select the optimal option for each realization.

Then, in (the beginning of) Year 9, the NPVs of these optimal options in Year 10 and the option of "continuing only with primary recovery" in Year 9 are again regressed on the production data from Year 1 to Year 8.

The same workflow is implemented until the time becomes Year 1. Other techniques can also be used in this context. Two other approaches, namely Gaussian Process Regression (GPR) and Support Vector Regression (SVR) are chosen in addition to Linear Regression. GPR and SVR will be briefed in the following section.

5. Machine learning techniques

5.1. Gaussian Process Regression (GPR)

GPR is a non-parametric ML approach that can be employed to perform data-driven modeling based on the Bayesian principle and Gaussian process. In a more technical sense, the Gaussian process (GP) can be perceived as a collection of random variables which possesses a multivariate joint distribution. In GPR, there is a function that can yield the output at certain inputs, in which the Gaussian noise with the normal distribution is included. GP acts as a distribution over functions and is defined by a mean function and covariance function. The covariance function (also known as kernel function) captures the dependence between different values of the function at their respective inputs. In this work, the employed kernel function is a squared exponential function. By having defined the mean and covariance functions, GP can be employed to retrieve a priori function values and posterior function values that are conditioned on the observed variables.

When it comes to the prediction of function values at new inputs, the joint distribution of the observed values and function values at these new points can be developed. Thereafter, GPR can be used to derive the "updated" posterior distribution by conditioning on these observed values. By doing so, the respective mean function can be determined by the posterior distribution and is treated as the prediction of regression. So, it is important to understand that GPR does not result in a deterministic model that best fits the data provided. However, it yields the predicted output by embracing the probability. For more comprehensive details of GPR, please counsel the following literature (Liu et al., 2020; Rasmussen and Williams, 2018). The modeling of GPR in this work is performed with the aid of Statistics and Machine Learning Toolbox in MATLAB R2021b (MathWorks, 2022). The hyperparameters are set at default values apart from the initial value for the noise standard deviation of GP which is set at 4.

5.2. Support Vector Regression (SVR)

SVR is another popular example of supervised learning techniques that is applied to approximate the relationship between inputs and the respective outputs with the weight vector and the bias term as the parameters. In general, SVR involves the mapping of the input space vector into feature space with higher dimensionality. This is to transform the initial non-linear problem into a more conveniently solvable linear regression function. Then, the regularized risk function can be minimized to estimate the weight vector and the bias term. To achieve this, the constrained optimization problem is established by introducing the non-negative slack variables (Forrester et al., 2008). This optimization function can be transformed into dual space by using Lagrange multipliers for the resolution of the constrained optimization problem (Shawe-Taylor and Cristianini, 2004). In this paper, the Gaussian function is used as the kernel function. Regarding the development of the SVR model, it is done by applying Statistics and Machine Learning Toolbox in MATLAB R2021b (MathWorks, 2022). The default hyperparameters are used, but standardization of data is implemented.

6. Results and discussion

Under the problem setting discussed, the DWOI consists of 2 years of primary recovery and then 8 years of waterflooding. This yields a field production of 10 years. DWOI corresponds to the alternative with the highest ENPV over all realizations. The respective EVWOI is 1479.06 million USD. This denotes that without acquiring any production data, the net profit considering all the realizations is 1479.06 million USD if there are 2 years of primary recovery followed by 8 years of water injection. Furthermore, DWPI corresponds to the alternative with the highest NPV for each realization. Averaging these NPVs results in EVWPI and it is calculated to be 1625.54 million USD. Then, VOPI is 146.47 million USD. This implies that if the cost of the information-gathering activity exceeds 146.47 million USD, this activity needs to be abandoned.

Also, the normalized frequency distributions (NFD) and the normalized cumulative frequency distributions (NCFD) of DWPI for the lifetime of primary recovery, those of secondary recovery (waterflooding), and a total lifetime of production are illustrated correspondingly in Fig. 2. Based on Fig. 2a, the NFD displays that about 30% of 50 geological realizations result in 1 and 2 years of primary recovery, which sums up to 60% of total realizations. On scrutiny, the NCFD portrays that 88% of the realizations recommend the lifetime of primary recovery to be equal to or less than 2 years. Thus, a considerably short lifetime of primary recovery is essential to achieve EVWPI. Additionally, about 36% of realizations yield 8 years of waterflooding as shown by the NFD plot in Fig. 2b. In this context, 70% of all the realizations propose water injection for at most 8 years. Around 56% of realizations proceed with a total of 10 years of production as demonstrated in Fig. 2c. In the case of NCFD, 44% of the realizations propose having a total lifetime of at most 9 years. It means that more than 50% of the realizations suggest 10 years of total lifetime.

Regarding the DWII of waterflooding for each realization, it has been previously expounded that 3 different techniques are employed to perform the backward induction (regression analysis) in the modified LSM algorithm to provide an SRDM solution. Out of these 3 techniques, GPR and SVR are generally considered ML-based. The regression analysis in the modified LSM algorithm can be treated as an example of a training process for ML techniques. Therefore, to elude the issue of overfitting during the training process, 5-fold cross-validation is used during regression analysis. Fig. 3 portrays the plots of observed NPV against approximated NPV for each alternative during regression analysis at each decision point in time with LR, GPR, and SVR.

As illustrated in Fig. 3, the Pearson Correlation Coefficients, ρ respectively obtained for LR, GPR, and SVR are 0.9634, 0.9772, and 0.9296. To further assess the quality of proximity, coefficient of determination, R^2 for LR, GPR, and SVR are correspondingly computed to be 0.9281, 0.9549, and 0.8642. According to these results, it is noticeable that GPR has outperformed both LR and SVR in terms of NPV approximation. On closer scrutiny, LR in the modified LSM algorithm yields an EVWII of 1490.58 million USD. GPR and SVR resulted in the corresponding EVWIIs of 1490.23 million USD and 1491.52 million USD. Moreover, VOIs of LR, GPR, and SVR are respectively 11.52 million USD, 11.17 million USD, and 12.46 million USD. Despite having the highest R^2, GPR results in the lowest VOI in this case. This shows that the higher accuracy of the approximated NPV (vs. observed NPV) for each alternative is unable to avoid the suboptimality of alternatives at certain paths (or realizations). Besides that, as compared with the case of LR, SVR enhances the VOI estimation by 8.18%. This reflects a good potential for VOI enhancement by implementing a nonlinear method under the framework of LSM.

Besides that, it can be deduced from these results that EVWIIs are higher than EVWOIs and this implies that it is worthwhile to include the effect of future information in decision-making. In this case, applying LR in the modified LSM algorithm would enhance the ENPV by 0.78%. An increase by 0.76% and 0.84% is also attained through the implementation of GPR and SVR, respectively. This practically illustrates the benefit of acquiring additional data in lieu of abiding by the initial plan as suggested by DWOI. Moreover, the NFD and NCFD of DWII for the lifetime of primary recovery, those of secondary recovery (waterflooding), and the total lifetime of production are illustrated correspondingly in Fig. 4 for LR, Fig. 5 for GPR, and Fig. 6 for SVR. Based on

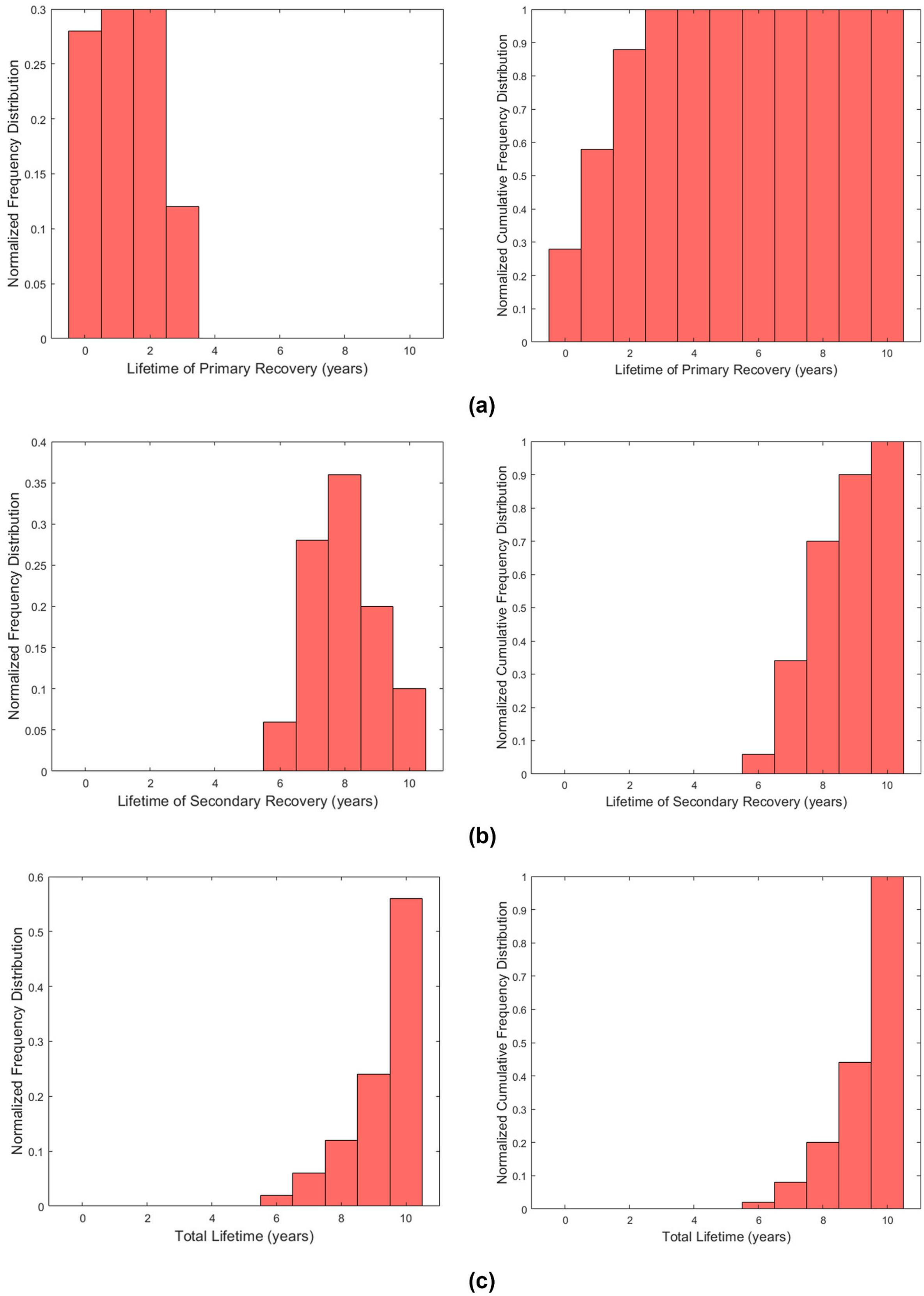

Fig. 2. Distribution of DWPI for: (a) Lifetime of Primary Recovery. (b) Lifetime of Secondary Recovery (Waterflooding). (c) Total Lifetime.

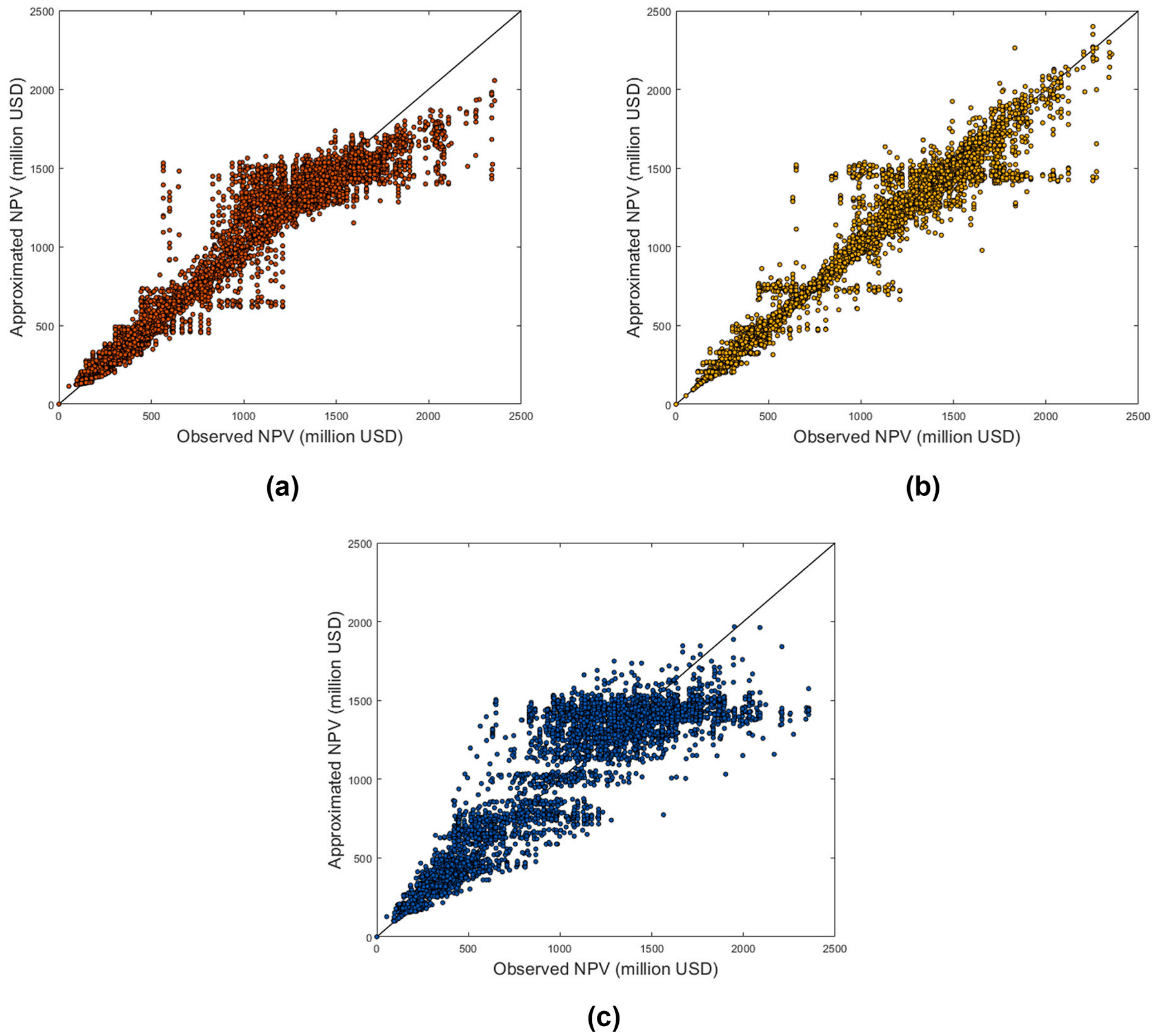

Fig. 3. Plots of observed NPV against approximated NPV for each alternative with (a) LR, (b) GPR, and (c) SVR.

the NCFD plots, it can be observed that only 20% of the realizations result in at most 1 year of primary recovery in the case of LR. Thus, there is 80% chance that 2 years of primary recovery produces the optimal results. For the lifetime of secondary recovery, 90% of all the realizations propose a water injection duration of at most 8 years. This results in 66% chance of 10 years of total lifetime.

Besides that, in the case of GPR, the NCFD plot illustrates that 36% of the realizations recommend the primary recovery of at most 1 year. In addition, there is 84% chance that the water injection should take place for at most 8 years. Regarding the total lifetime, GPR results in a total period of 10 years with 58% chance. When it comes to the NCFD plot of SVR, 98% of the realizations result in primary recovery for at most 2 years whereas 94% of those suggest a waterflooding of at most 8 years. Thereby, 62% of all the realizations result in 10 years of total lifetime. In general, these three techniques (LR, GPR, and SVR) would mostly result in the optimal decision of 2 years of primary recovery and 8 years of waterflooding that contribute to a total of 10 years of production. As it has been explained, DWPIs as suggested in Fig. 2 signify the most optimal decisions. So, if the distribution of DWII is closer to that of DWPI, there is a better chance for the respective EVWII to be higher. Nonetheless, the distributions of DWII for the Lifetime of Primary Recovery estimated by using LR, GPR, and SVR are considerably different from that of DWPI. This also explains the obvious difference between each of the EWIIs and EWPI.

Fig. 7 compares the cumulative distribution function (CDF) of the NPVs corresponding to DWOI, DWII (considering all 3 techniques), and DWPI. According to Fig. 7, the more rightward the CDF curve is, the higher the ENPV is. The CDF of NPV_{DWOI} and the three CDFs of NPV_{DWII} are close to each other. This proximity resonates with the slight improvement (less than 1%) in the EVWOI for the determination of VOI, as discussed earlier. This can be due to the suboptimality of alternatives made for some realizations as the ML-based regressions used are approximate methods.

Fig. 8 (Fig. 9) shows the plot of the mean oil (water) production rate corresponding to DWOI and DWII of 3 different techniques. In the case of DWOI, the mean oil production rate starts increasing after Year 2 because waterflooding is initiated at that time. This is also reflected by the increase in the mean water production rate after Year 2 as shown in Fig. 9. For DWIIs of the three techniques, the initiation time of waterflooding is generally different for different realizations based on the acquisition of information under the framework of SRDM. In this aspect, the issue of suboptimality, as discussed earlier, would occur, and affect the trends of the plots. A tremendous increase after Year 2 is observed. This can be explained by referring to Figs. 4–6, from which more than 50% of the NFD (optimal decision policy) correspond to the lifetime of primary recovery for 2 years.

Despite being limited by the curse of dimensionality due to the increase in the number of alternatives, this work successfully displays the integration of NRS into the framework of modified LSM for optimization of the initiation time of waterflooding under uncertainties. Different

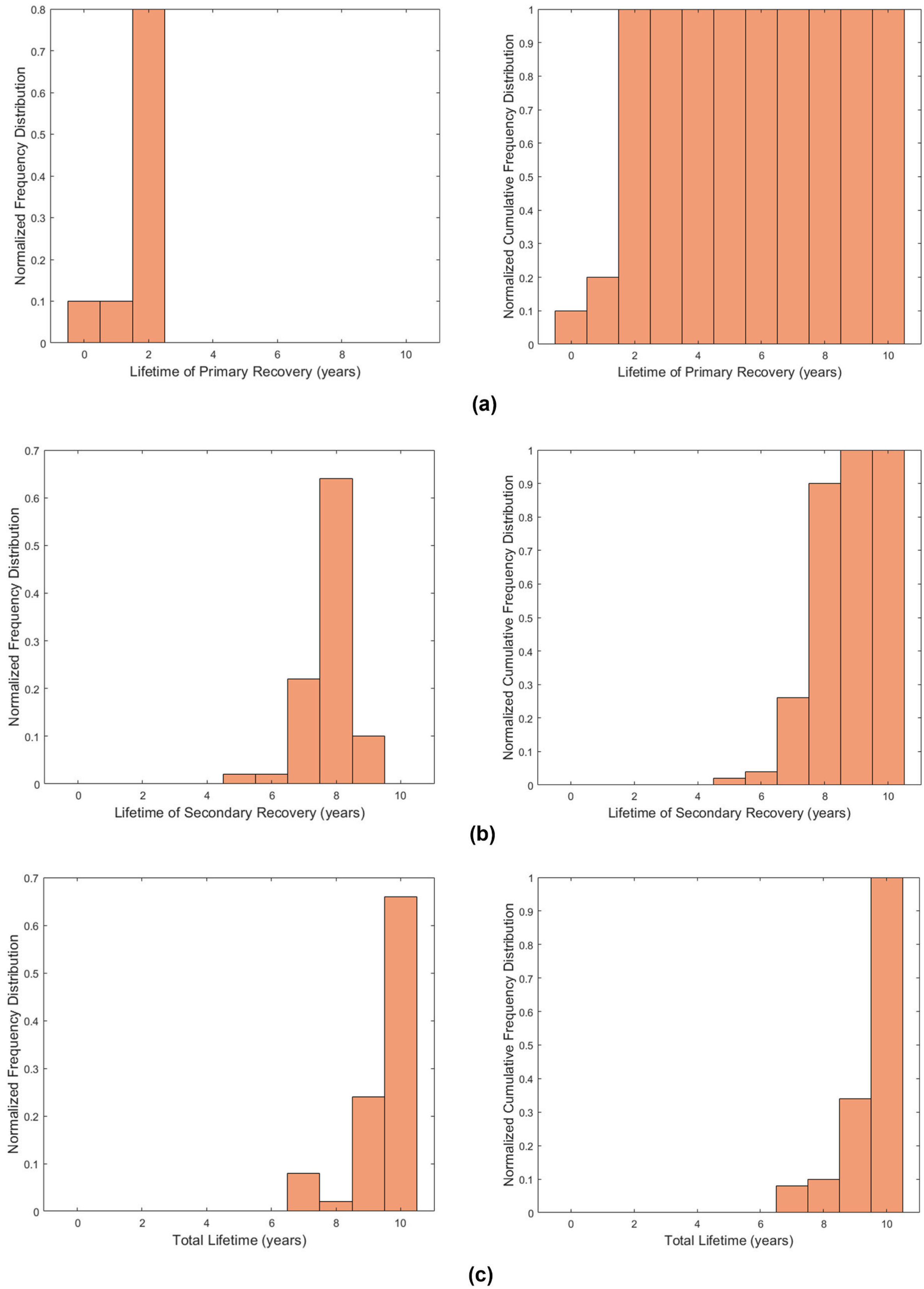

Fig. 4. Distribution of DWII for lifetime of primary recovery, lifetime of secondary recovery (waterflooding), and total lifetime in LR.

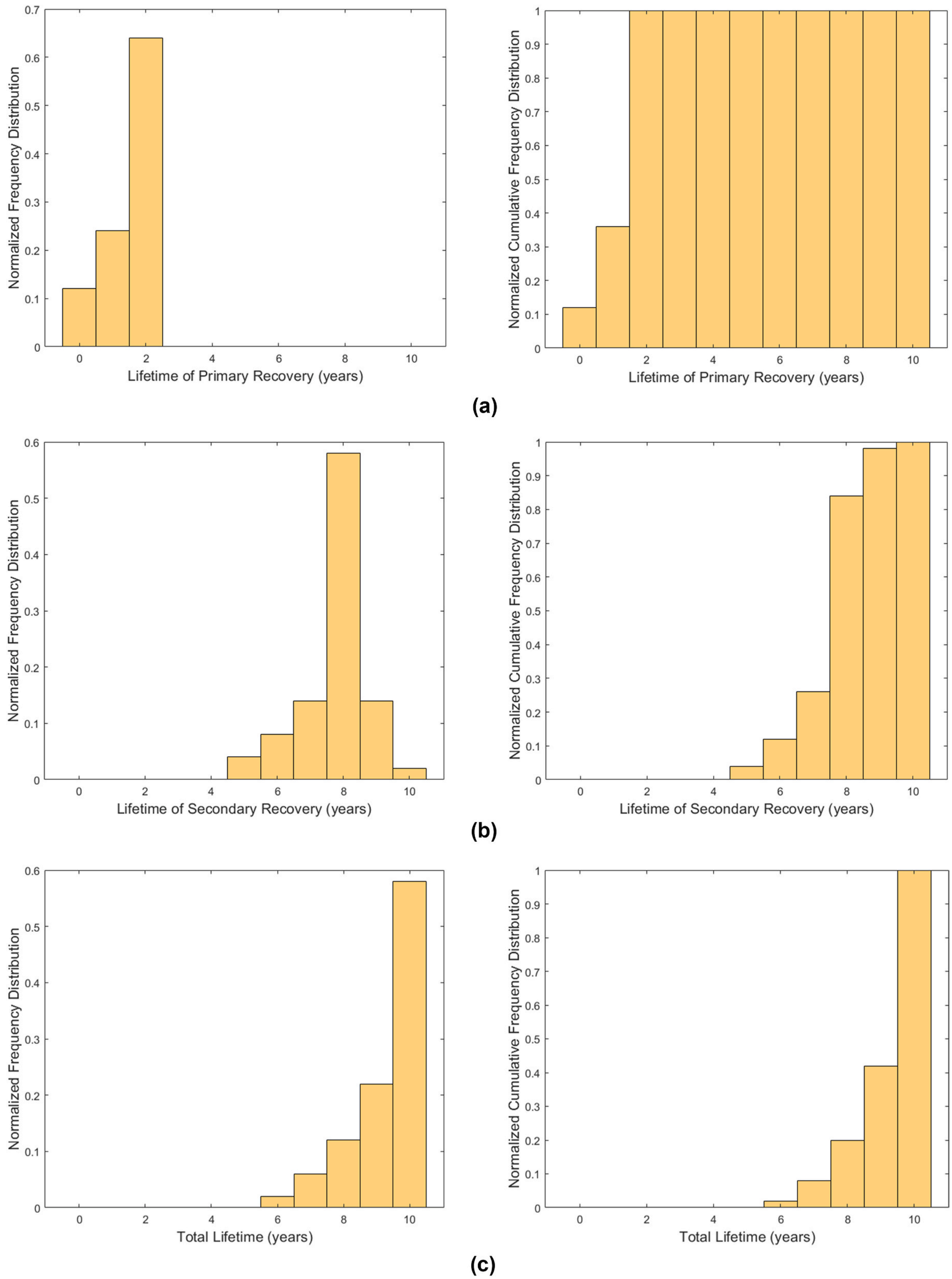

Fig. 5. Distribution of DWII for lifetime of primary recovery, lifetime of secondary recovery (waterflooding), and total lifetime in GPR.

geological uncertainties are considered as previously mentioned to increase the verisimilitude of the case study. Including other types of uncertainty, such as economic uncertainty, is another viable part of future works that can further reinforce the practicality of this methodology. In that case, prices or costs can be modeled by employing a stochastic approach, viz. Two-Factor Price Model (Jafarizadeh and Bratvold, 2013). Nevertheless, for practical purposes, decision makers need to honor the trade-off between uncertainties and the availability of resources (e.g., financial or labor). This is to ensure that limited resources will not be exhausted to include as many uncertainties as possible.

Some possible extensive applications can be considered in the future

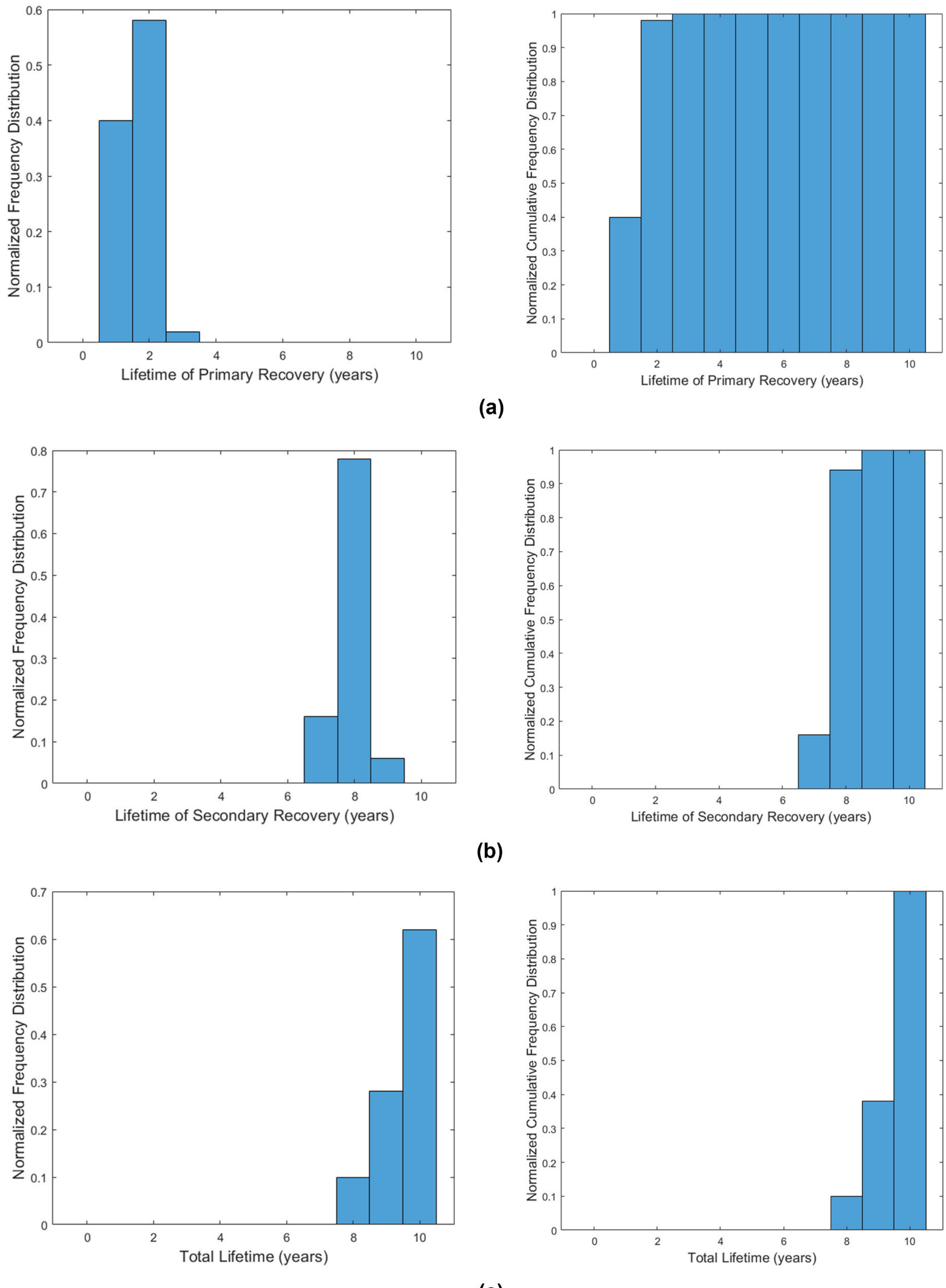

Fig. 6. Distribution of DWII for lifetime of primary recovery, lifetime of secondary recovery (waterflooding), and total lifetime in SVR.

for the work presented here. One of them includes the application of smart proxy models (Mohaghegh, 2022), as substitutes for NRS, under the framework of the modified LSM algorithm. In general, NRS is considered one of the prevalent tools when it comes to RM issues on a field scale. Nonetheless, using NRS, a geologically complex reservoir model is likely to be computationally prohibitive to be coupled with the modified LSM algorithm. Concerning this, the application of smart proxy models in tandem with the algorithm can be a good recommendation to

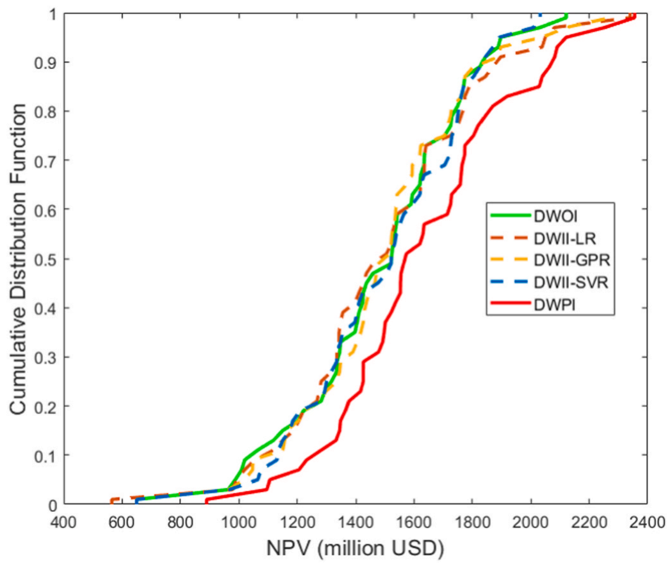

Fig. 7. CDF of NPVs corresponding to DWOI, DWII (considering all techniques), and DWPI.

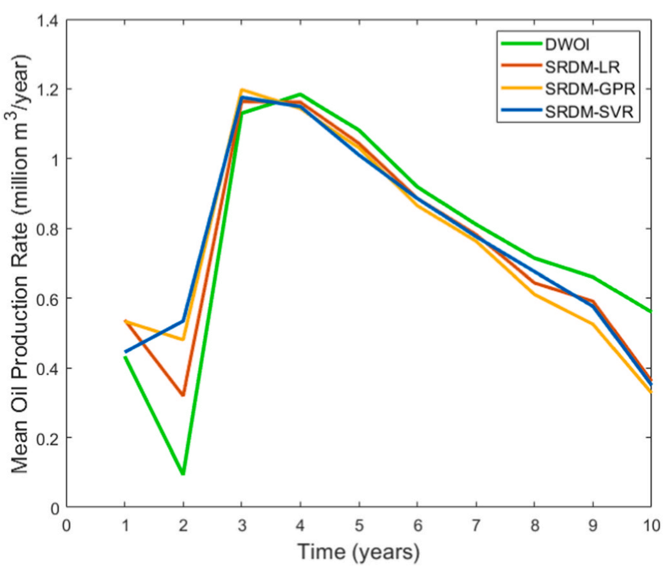

Fig. 8. Plot of mean oil production rate corresponding to DWOI and DWII of all three techniques.

be considered in the future. This demonstrates not only the versatility of the algorithm but also reduces the computational efforts induced by NRS under the paradigm of the SRDM. In this context, the good computational reduction capability of the smart proxy models has been demonstrated in several pieces of literature (Nait Amar et al., 2020, 2021; Ng et al., 2021a, 2021b, 2022).

This work mainly sheds light on the use of supervised learning in the context of the SRDM framework. The robustness of ML can be further highlighted if the use of a more advanced technique, namely reinforcement learning (RL), is embedded in this framework. RL (van Otterlo and Wiering, 2012), generally expounds on the interaction between an intelligent model (an agent) and an environment (a problem setting) to take actions based on the reward. In other words, RL can be perceived as a DM tool. It is thereby worth investigating how RL can be combined with the LSM algorithm for wider applications to improve DM in the aspects of RM. Additionally, the SRDM approach discussed here can also be extended to other domains of reservoir and production engineering,

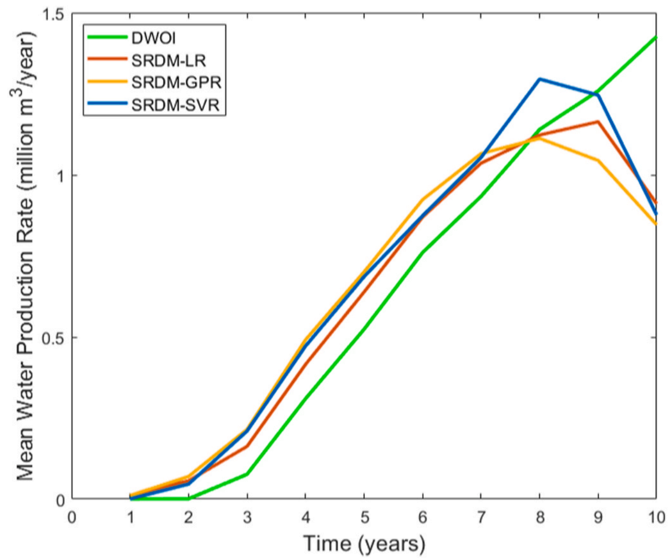

Fig. 9. Plot of mean water production rate corresponding to DWOI and DWII of all three techniques.

such as water-alternating-gas injection, well-placement optimization, and inflow control valve optimization. These optimization problems, to certain extents, are perceived as an example of switching problems, which have been proven to be efficiently resolved by employing the modified LSM algorithm.

7. Summary and conclusions

In this work, it has been illustrated and expounded on how the modified LSM algorithm, an epitome of the simulation-regression approach, can be implemented as an SRDM approach to resolve a sequential decision problem in reservoir engineering. Concerning this, optimization of the initiation time of waterflooding in the OLYMPUS reservoir model under geological uncertainties is chosen as the pertinent sequential decision problem. Being different from the initial LSM algorithm proposed by Longstaff and Schwartz (2001), this modified variant integrates the dependency on previously and currently acquired data. In this aspect, the effect of information is shown to be integrated into the context of DM. To enlighten the readers, the mathematical formulations to compute VOI have been concisely explained. There is also a discussion and illustration of how the modified LSM algorithm (as a variant of the simulation-regression approach) can play a part in determining VOI, which is one of the most prevalent DM tools. Besides that, it has been discussed how this algorithm can be implemented as an SRDM approach to resolve the issue of waterflooding initiation time. LR has been the conventional technique of the LSM algorithms. Apart from LR, two other ML-based techniques, viz. GPR and SVR are employed to conduct the regression analysis. Based on our investigation, the DWOI is 2 years of primary recovery followed by 8 years of waterflooding, and the resulting EVWOI is 1479.06 million USD. With the aid of the SRDM approach, the VOIs that are correspondingly estimated by using LR, GPR, and SVR are 11.52 million USD, 11.17 million USD, and 12.46 million USD. Thereafter, the EVWIIs which are estimated by LR, GPR, and SVR, correspond to 1490.58 million USD, 1490.23 million USD, and 1491.52 million USD, respectively. Thus, SVR improves ENPV by the highest percentage, which is 0.84% despite displaying the lowest accuracy during regression analysis. VOI that is approximated by GPR (with the highest accuracy of regression analysis) shows a slightly inferior result, that is improvement of the ENPV by 0.76%. This can be generally explained by the sub-optimality of decisions due to approximation error. Nevertheless, SVR illustrates an improvement of the estimated VOI.

Albeit it is demonstrated that employing non-linear regression ML-based techniques does not guarantee an improvement of VOI in this work (as compared with the VOI approximated by using LR), it provides an insightful demonstration regarding the application of these ML techniques in the context of VOI determination. Also, applying this SRDM approach to the OLYMPUS model can serve as a step closer to the resolution of real-world sequential decision problems. Despite the positive results garnered from this study, several limitations, especially on computational cost due to the forward modeling of a more sophisticated reservoir and the higher number of alternatives, are to be addressed to further improve this methodology. Uncertainty modeling of prices and integration with RL are also considered to improve the robustness of this methodology in the future.

Credit author statement

Cuthbert Shang Wui Ng: Data curation, Formal analysis, Methodology, Investigation and Modeling, Coding and Programming, Writing, Editing. Ashkan Jahanbani Ghahfarokhi: Supervision, Writing, Reviewing and Editing.

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.

Data availability

The authors are unable or have chosen not to specify which data has been used.

Acknowledgment

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). The authors acknowledge the suggestions, ideas, and insights provided by Dr. Reidar Brumer Bratvold and Dr. Aojie Hong in the early phase of the formulation of this work. The authors also thank Mr. Wilson Wiranda from the Department of Geoscience and Petroleum, NTNU, for the preparation of the OLYMPUS data.

References

Alkhatib, A., Babaei, M., King, P.R., 2013. Decision making under uncertainty: applying the least-squares Monte Carlo method in surfactant-flooding implementation. SPE J.
Anyosa, S., Bunting, S., Eidsvik, J., Romdhane, A., Bergmo, P., 2021. Assessing the value of seismic monitoring of CO2 storage using simulations and statistical analysis. Int. J. Greenh. Gas Control 105.
Bratvold, R.B., Begg, S.H., 2010. Making Good Decisions, first. ed. Society of Petroleum Engineers, Richardson, Texas.
Bratvold, R.B., Bickel, J.E., Lohne, H.P., 2009. Value of information in the oil and gas industry: past, present and future. SPE Reservoir Eval. Eng. 12.
Dutta, G., Mukerji, T., Eidsvik, J., 2019a. Value of information analysis for subsurface energy resources applications. Appl. Energy 252.
Dutta, G., Mukerji, T., Eidsvik, J., 2019b. Value of information of time-lapse seismic data by simulation-regression: comparison with double-loop Monte Carlo. Comput. Geosci. 23.
Eidsvik, J., Dutta, G., Mukerji, T., Bhattacharjya, D., 2017. Simulation–regression approximations for value of information analysis of geophysical data. Math. Geosci. 49.
Fonseca, R.M., Rossa, E. Della, Emerick, A.A., Hanea, R.G., Jansen, J.D., 2020. Introduction to the special issue: overview of OLYMPUS optimization benchmark challenge. Comput. Geosci.
Forrester, A.I.J., Sóbester, A., Keane, A.J., 2008. Engineering Design via Surrogate Modelling : a Practical Guide. J. Wiley.
Hillson, D., Murray-Webster, R., 2017. Understanding and managing risk attitude. In: Second Edition, Understanding and Managing Risk Attitude, second ed.
Hong, A., Bratvold, R.B., Lake, L.W., 2019. Fast analysis of optimal improved-oil-recovery switch time using a two-factor production model and least-squares Monte Carlo algorithm. SPE Reservoir Eval. Eng.
Hong, A.J., Bratvold, R.B., Thomas, P., Hanea, R.G., 2018. Value-of-information for model parameter updating through history matching. J. Pet. Sci. Eng. 165.
Howard, R.A., 1980. An assessment of decision analysis. Oper. Res. 28, 4–27.
Howard, R.A., 1966. Information value theory. IEEE Trans. Syst. Sci. Cybern. 2.
Howard, R.A., Abbas, A.E., 2016. Foundations of Decision Analysis. Pearson Education Limited, Harlow, England.
Jafarizadeh, B., Bratvold, R.B., 2013. Sell spot or sell forward? Analysis of oil-trading decisions with the two-factor price model and simulation. SPE Econ 5. Manag.
Liu, H., Ong, Y.S., Shen, X., Cai, J., 2020. When Gaussian process meets big data: a review of scalable GPs. IEEE Transact. Neural Networks Learn. Syst. 31.
Longstaff, F.A., Schwartz, E.S., 2001. Valuing American options by simulation: a simple least-squares approach. Rev. Financ. Stud. 14.
MathWorks, T., 2022. MATLAB (R2021b). MathWorks Inc.
Mohaghegh, S.D., 2022. Smart Proxy Modeling: Artificial Intelligence and Machine Learning in Numerical Simulation. CRC Press, Boca Raton.
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.
Nait Amar, M., Zeraibi, N., Jahanbani Ghahfarokhi, A., 2020. Applying hybrid support vector regression and genetic algorithm to water alternating CO2 gas EOR. Greenh. Gases Sci. Technol.
Ng, C.S.W., 2019. Using the Least-Squares Monte Carlo Algorithm to Optimize IOR Initiation Time. Norwegian University of Science and Technology, Trondheim.
Ng, C.S.W., Ghahfarokhi, A.J., Nait Amar, M., 2022. Production Optimization under Waterflooding with Long Short-Term Memory and Metaheuristic Algorithm. Petroleum.
Ng, C.S.W., Jahanbani Ghahfarokhi, A., Nait Amar, M., 2021a. Application of nature-inspired algorithms and artificial neural network in waterflooding well control optimization. J. Pet. Explor. Prod. Technol.
Ng, C.S.W., Jahanbani Ghahfarokhi, A., Nait Amar, M., Torsæter, O., 2021b. Smart proxy modeling of a fractured reservoir model for production optimization: implementation of metaheuristic algorithm and probabilistic application. Nat. Resour. Res. 30, 2431–2462.
Olson, R.S., Bartley, N., Urbanowicz, R.J., Moore, J.H., 2016. Evaluation of a tree-based pipeline optimization tool for automating data science. In: GECCO 2016 - Proceedings of the 2016 Genetic and Evolutionary Computation Conference.
Parra Sanchez, C., 2010. A Life Cycle Optimization Approach to Hydrocarbon Recovery. The University of Texas at Austin, Austin.
Powell, W.B., 2011. Approximate dynamic programming: solving the curses of dimensionality. In: Second Edition, Approximate Dynamic Programming: Solving the Curses of Dimensionality, second ed.
Rasmussen, C.E., Williams, C.K.I., 2018. Gaussian Processes for Machine Learning, Gaussian Processes for Machine Learning.
Satter, A., Varnon, J.E., Hoang, M.T., 1998. Integrated reservoir management. SPE Repr. Ser.
Schlumberger, 2019. Eclipse Reservoir Simulation Software Reference Manual. schlumberger.
Shawe-Taylor, J., Cristianini, N., 2004. Kernel Methods for Pattern Analysis, Kernel Methods for Pattern Analysis.
Tadjer, A., Bratvold, R.B., Hong, A., Hanea, R., 2021a. Application of machine learning to assess the value of information in polymer flooding. Pet. Res. 6.
Tadjer, A., Hong, A., Bratvold, R.B., 2021b. A sequential decision and data analytics framework for maximizing value and reliability of CO2 storage monitoring. J. Nat. Gas Sci. Eng. 96, 104298.
van Otterlo, M., Wiering, M., 2012. In: Wiering, M., van Otterlo, M. (Eds.), Reinforcement Learning and Markov Decision Processes BT - Reinforcement Learning: State-Of-The-Art. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 3–42.
Wiggins, M.L., Startzman, R.A., 1998. An approach to reservoir management. SPE Repr. Ser.
Willigers, B.J.A., Begg, S.H., Bratvold, R., 2011. Valuation of swing contracts by least-squares Monte Carlo simulation. In: SPE Economics and Management.
Willigers, B.J.A., Bratvold, R.B., 2009. Valuing oil and gas options by least-squares Monte Carlo simulation. In: SPE Projects, Facilities and Construction.