processes

MDPI

Article

A Novel Surrogate-Assisted Multi-Objective Well Control Parameter Optimization Method Based on Selective Ensembles

Lian Wang [1,2,3,*], Rui Deng [1,3], Liang Zhang [1,3], Jianhua Qu [1,3,*], Hehua Wang [1,3], Liehui Zhang [2], Xing Zhao [1,3], Bing Xu [1,3], Xindong Lv [3,4] and Caspar Daniel Adenutsi [5]

1 Zhenhua Oil Co., Ltd., Beijing 100031, China; dengrui@zhenhuaoil.com (R.D.); zhangliang@zhenhuaoil.com (L.Z.)
2 State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation, Southwest Petroleum University, Chengdu 610500, China
3 Chengdu North Petroleum Exploration and Development Technology Co., Ltd., Chengdu 610051, China
4 College of Energy, Chengdu University of Technology, Chengdu 610059, China
5 Reservoir Simulation Laboratory, Department of Petroleum Engineering, Faculty of Civil and Geo-Engineering, Kwame Nkrumah University of Science and Technology, Kumasi AK-385-1973, Ghana; casdanad@hotmail.com
* Correspondence: wanglian02@163.com (L.W.); qujianhua@zhenhuaoil.con (J.Q.)

Abstract: Multi-objective optimization algorithms are crucial for addressing real-world problems, particularly with regard to optimizing well control parameters, which are often computationally expensive due to their reliance on numerical simulations. Surrogate-assisted models help to reduce this computational burden, but their effectiveness depends on the quality of the surrogates, which can be affected by candidate dimension and noise. This study proposes a novel surrogate-assisted multi-objective optimization framework (MOO-SESA) that combines selective ensemble support-vector regression with NSGA-II. The framework's uniqueness lies in its adaptive selection of a diverse subset of surrogates, established prior to iteration, to enhance accuracy, robustness, and computational efficiency. To our knowledge, this is the first instance in which selective ensemble techniques with multi-objective optimization have been applied to reservoir well control problems. Through employing an ensemble strategy for improving the quality of the surrogate model, MOO-SESA demonstrated superior well control scenarios and faster convergence compared to traditional surrogate-assisted models when applied to the SPE10 and Egg reservoir models.

Keywords: selective surrogate ensembles; multi-objective optimization; support vector regression; NSGA-II

check for updates

Citation: Wang, L.; Deng, R.; Zhang, L.; Qu, J.; Wang, H.; Zhang, L.; Zhao, X.; Xu, B.; Lv, X.; Adenutsi, C.D. A Novel Surrogate-Assisted Multi-Objective Well Control Parameter Optimization Method Based on Selective Ensembles. *Processes* **2024**, *12*, 2140. https://doi.org/10.3390/pr12102140

Academic Editor: Jose Carlos Pinto

Received: 25 August 2024
Revised: 24 September 2024
Accepted: 26 September 2024
Published: 1 October 2024

Copyright: © 2024 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

One of the most essential steps prior to production and after drilling and completion is numerical simulation-based well control optimization. This refers to finding an optimum scheme that could achieve higher productivity and profitability [1,2]. The net present value (NPV), which displays the comprehensive performance of developing the reservoir, is usually considered as the single objective for optimizing well control schemes [3]. However, oilfield development is complicated, as well control optimization problems are generally related to multiple objectives, including optimizing profitability and oil recovery while reducing operation fees and investments [4]. Moreover, these well control-related objectives may conflict with each other [5]. Two main methods, namely decomposition and the direct method, were utilized to search for the optimal solution of multi-objective problems. The drawback of the decomposition method, such as weighted aggregation, is the selection of the correct weight setting. In comparison, when the direct method is used, the optimum is reached with no human interference. By considering all objectives simultaneously, multi-objective optimization provides a compromise solution set for decision-makers. Individuals in the compromise solution set are non-dominated to each other, and together, they comprise the

Pareto front (PF) [6]. To that end, multi-objective optimization, which simultaneously takes into account more than one objective, is more suitable for well control optimization [7].

In previous studies on addressing well control multi-objective optimization problems, many researchers have paid attention to two main techniques [8,9], namely gradient-based and gradient-free techniques, to obtain high-quality solutions that are represented by diversity and convergence. The gradient-based strategy is one of the traditional techniques, and it is also perfect in the field of computational efficiency during the process of numerical simulation [10]. Within the realm of gradient-based methods, the adjoint strategy, which relies on the simulation model and the reverse adjoint equation, results in neglecting the dimension of optimal variables. In this way, the adjoint technique is an efficient way to address well control optimization problems with practical reservoirs. Nevertheless, obtaining the adjoint gradient requires revising the code of simulation software to acquire the Jacobian matrix, which limits the application of the adjoint method [11,12]. Furthermore, the adjoint method has the limitation of falling into the local optimal while missing the global optimal [13]. Meanwhile, the gradient-free algorithm has also been extensively utilized in multi-objective reservoir optimization problems in recent years because of its broad suitability [14,15]. Specifically, Yasari et al. (2013) presented a robust multi-objective well control parameters optimization framework that considered three specified NPVs as objectives [16]. Luo et al. (2015) established a multi-objective well control parameters optimization framework with the particle swarm optimization (PSO) algorithm and estimation of distribution algorithm [17]. Fu and Wen (2017) compared the multi-objective particle swarm optimization and PSO in production optimization problems [18]. However, evaluating a lot of objectives to search for the optimum solution increased the computational burden and reduced the calculation efficiency. These studies have achieved impressive performance in the optimization of well control parameters, but the simulations that run during the optimization process are time-consuming. According to the grid size of the numerical model, it takes hours to run the simulation process. A natural concern that emerges is how to simultaneously reduce the computational load and maintain accuracy.

To lighten the computational cost and increase the optimization efficiencies during the well control optimization process, Proper Orthogonal Decomposition (POD), data-driven model and machine learning-based surrogate methods are the most employed methods. POD is brief in its computation, simple to execute, and offers a perfect approximation for complex numerical simulation systems [19]. Using the POD model, the optimization process could be more efficient. Similar to the traditional gradient technique, POD also requires the simulator code to be modified, which is difficult for a commercial simulator. Without any prior knowledge of geological and numerical models [20], a data-driven model could serve as a surrogate for well control optimization based on history matching of production data [21]. There are three main types of data-driven models, and these include the capacitance-resistance model, interwell numerical simulation model, and its front tracking [22–24]. Meanwhile, the main challenge associated with the data-driven model is that a lot of production data are required. As one of the most widely used approaches, machine learning theory-based surrogate methods can flexibly model any inputs and outputs relationship [2] with enough training samples, such that the expensive computation of the simulator could be lessened during the optimization process. Many machine learning approximation methods, such as the Gaussian process [25], artificial neural network [26], convolutional neural networks [27], support vector regression [23,28], and radial basis function [29], have been successfully employed in establishing surrogate models for computationally expensive problems. However, these are single or double models applied as surrogates, such that the quality of the model is difficult to guarantee. Meanwhile, the accuracy of machine learning surrogates is highly affected by the number of optimal variables and the scale of the training samples. With that in mind, the first task is to enhance the quality and robustness of the models with limited training samples. It is worth mentioning that the surrogate ensemble strategy, which constructs several single

surrogate models, is suitable for improving the quality of the optimal solutions and has not been utilized in the field of solving well control optimization problems.

The goal of this study is to propose a novel surrogate model-based well control optimization framework with machine learning and intelligent algorithm, which could make full use of training samples to guide the multi-objective optimization search. To that end, several support vector regression (SVR) models are established before iteration and a simple, yet diverse, subset is adaptively selected, based on ensemble learning during the optimization process, to develop the surrogate ensemble while reducing the computational cost. More importantly, this is the first application of selective ensemble techniques with multi-objective optimization to reservoir well control problems.

2. The Multi-Objective Well Control Optimization Model

In well control optimization, the model typically involves identifying optimal variables and defining objective functions to improve production efficiency and profitability. In this study, the optimal variables include key well control parameters such as injection rates and bottom hole pressures, which directly influence reservoir performance. To assess the effectiveness of these strategies, two main objective functions are established. The first is NPV, a classic economic indicator used to evaluate the profitability of production by accounting for revenues, operating costs, and discounting future income. NPV serves as the primary objective in this study, aiming to maximize the financial returns from the reservoir. The second objective is cumulative oil production (COP), which measures the total volume of oil extracted and focuses on maximizing recovery efficiency. Because NPV and COP can sometimes conflict—where maximizing one may impact the other—this is a multi-objective optimization problem. The goal is to explore trade-offs between profitability and recovery, ultimately finding a balance that ensures both economic success and efficient reservoir depletion. The relationships between decision variables and objectives are formed as follows [2]:

$$J_1(u) = \sum_{n=1}^{N_t} \frac{\Delta t_n}{(1+d)^{\frac{t_n}{365}}} \left[\sum_{j=1}^{P} \left(r_o \cdot q_{o,j}^n - r_w \cdot q_{w,j}^n \right) - \sum_{j=1}^{I} \left(c_{wi} \cdot q_{wi,j}^n \right) \right] \quad (1)$$

$$J_2(u) = \sum_{n=1}^{N_t} \sum_{j=1}^{P} q_{o,j}^n \quad (2)$$

The specific meanings of each symbol of Equations (1) and (2) are listed in the Nomenclature Section.

3. Related Method and Algorithm

3.1. Selective Surrogate Ensembles

Ensemble learning, as a surrogate management technique for machine learning, constructs a lot of base models and combines them as a strong model with more accuracy and robustness than a single model [30]. The two mainstream ensemble methods, namely boosting and bagging, have different strategies. Bagging involves minimizing variance using a parallel ensemble while boosting is minimizing bias by a sequential ensemble [31]. Moreover, during the iteration process, the bias of surrogates is less important. For this reason, bagging was adopted in this study to establish the ensemble. Meanwhile, subset learners were selected to improve the quality of the surrogate ensemble. Figure 1 consists of four steps that show the workflow of selective bagging.

As can be seen in Figure 1, each datapoint in N subsets (S_1, S_2, …, S_N) is randomly selected by a bootstrap that resamples the initial dataset. The traditional way of deciding the size of each subset is half-sampling. To make subsets more representative, a probability-based sampling strategy was employed in this study. Each point in the initial dataset was set to have a probability of 50% to be selected for each subset. This means the size of subsets is not fixed, which promotes the diversity of the ensembles. Then, the N SVR models were

trained based on subsets S_1, S_2, . . ., and S_N separately. The specific process of establishing the surrogate ensemble is elaborated in Algorithm 1.

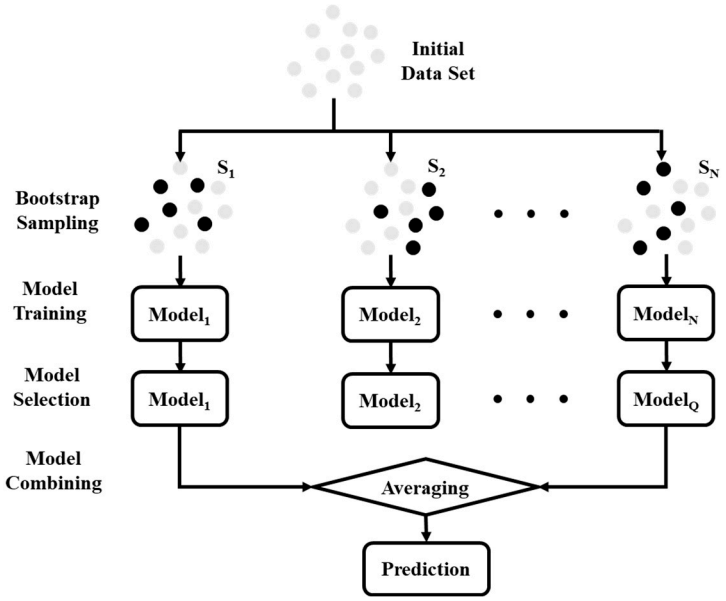


Figure 1. The main steps of selective bagging (The gray dots are the initial data samples, and the black dots are the selected data samples).

Algorithm 1 Pseudocode of establishing the surrogate ensemble

Input: *D*-the dataset, *N*-the number of models
for *i* from 1 to *N*
  Empty S_i
  for individual in *D*
    if $U(0,1) < 50\%$
      Add the individual to S_i
    end if
  end for
end for
for *i* from 1 to *N*
  Establish an SVR model M_i using S_i
end for
Output: subset dataset (S_1, S_2, . . ., S_N) and model set (M_1, M_2,. . ., M_N)

Another challenge related to selective ensembles relates to the type and method employed to select the SVR models for combinations. Within limits, the accuracy of surrogate ensemble improves with the increase in the selective ensemble size [32]. This indicates that the model selection strategy is essential for reducing the size of the ensemble without decreasing its quality. To enhance the accuracy and robustness of the surrogate ensemble, a random selection strategy was proposed for selecting SVR models in each generation. More specifically, fixed Q models were selected from the N models in each generation. Choosing the best individual (g_b) in the current generation, fitness values (F_1, F_2, . . ., F_N) of g_b with N models (M_1, M_2, . . ., M_N) were evaluated. Subsequently, the N SVR models were sorted based on the fitness values and separated into equal Q groups. To this end, one SVR for each data group was randomly selected and the surrogate ensemble for the next iteration was established. Algorithm 2 details the model selection strategy of the selective ensemble.

A model set with eight models (M_1, M_2, . . ., M_8) is taken as an example to display the principle of selection strategy. Let the fitness value of g_b be $F_1 = 1.1$, $F_2 = 1.4$, $F_3 = 1.6$, $F_4 = 1.5$, $F_5 = 1.7$, $F_6 = 1.8$, $F_7 = 1.2$, $F_8 = 1.3$, respectively. To construct a surrogate ensemble

with four models for the next iteration, these models are ranked according to fitness values and divided into four groups (F_1, F_7), (F_8, F_2), (F_4, F_3), and (F_5, F_6). Then, one model is chosen at random from every group. This way, F_1, F_8, F_4, and F_6 could be an output case.

Algorithm 2 Pseudocode of model selection strategy

Input: Q-the scale of ensemble after selection, g_b-the best individual of current generation, (M_1, M_2,. . ., M_N)-the model set
1: **if** initial generation
2: Select Q models randomly for model set
3: **else**
4: Calculate fitness values (F_1, F_2, . . ., F_N) of g_b with (M_1, M_2, . . ., M_N)
5: Sort N SVR models by (F_1, F_2, . . ., F_N)
6: Divide N sorted SVR into equal Q groups
7: **for** i = 1: Q
8: Randomly select a model from group i and establish surrogate model
9: **end for**
10: **end if**
Output: SVR models

3.2. Optimization Algorithm (NSGA-II)

NSGA-II, as the most traditional and perfect multi-objective optimization algorithm, was introduced by Deb [6]. Its workflow includes a specific sequence of cross, mutative, and non-dominated sorting. The key distinction of NSGA-II from other algorithms lies in its sorting strategy, where individuals are ranked based on crowding distance. Additionally, the elitism strategy applied in NSGA-II ensures that the best individuals are preserved for the next iteration, promoting faster convergence. Given its strong performance, NSGA-II was selected as the algorithm for this study.

3.3. Specific Workflow of MOO-SESA

The main framework of MOO-SESA, including two essential strategies, namely model selection and non-dominated sorting, is displayed in Figure 2. Firstly, the decision variables and objectives were determined as the initial conditions. Next, the initial individuals were produced by LHS and related objective values were calculated by the simulator to construct the dataset. Then, a selective ensemble SVR model was constructed to reduce the computational burden based on the dataset and ensemble strategy. This could make full use of limited samples and lead to more efficient and accurate surrogates. During each iteration, the surrogate ensemble was reconstructed. Finally, the NSGA-II was performed as the optimizer based on non-dominated sorting. The selective surrogate ensemble enhances the quality of prediction accuracy and releases the computational burden.

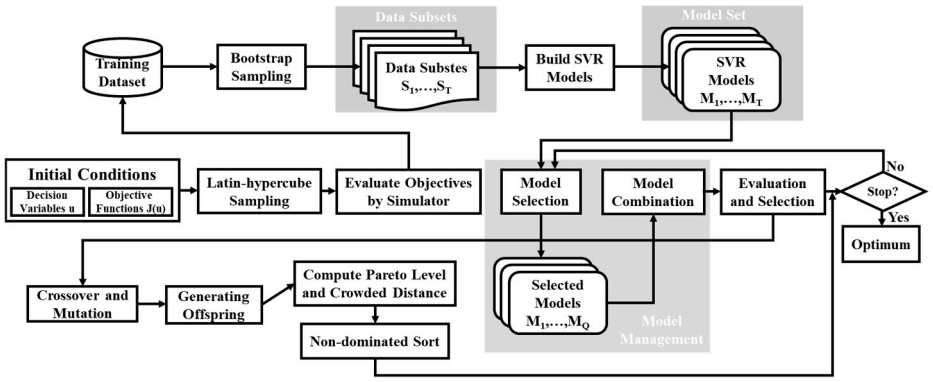


Figure 2. The specific workflow of MOO-SESA.

4. Case Study

To evaluate the performance of the MOO-SESA framework, two synthetic reservoir models are employed and the results are compared with those from the classical surrogate-based optimization framework that integrates a single static surrogate model with an optimization algorithm (SVR-NSGA-II). The first reservoir model was a two-dimensional extract from the SPE10 model, while the second was a three-dimensional Egg model based on Jansen et al. (2014) [33]. Simulations were conducted using the MATLAB Reservoir Simulation Toolbox (MRST) as described by Lie (2019) [34].

4.1. Case 1

To assess the performance of the MOO-SESA framework in this scenario, a two-dimensional heterogeneous reservoir model derived from a portion of layer four of the SPE10 model was utilized. This model is configured in a five-spot pattern, comprising one central injector and four producers located at the corners of the grid. The grid consists of 50 × 50 × 1 blocks, with each cell measuring 6 × 6 × 10 m. The arrangement places the injector at the center of the model, optimizing the injection and production process. The logarithmic permeability field for this setup is depicted in Figure 3.

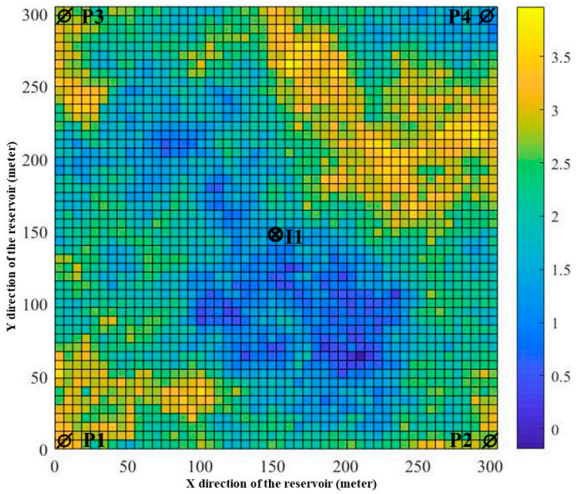


Figure 3. The specific permeability field of case 1 (P1–P4 are producers, I1 is injector).

In this case study, the decision variables are the injection rate of the injector and the bottom hole pressures (BHP) of the four producers. The injection rate ranges from 0 to 150 m^3/day, and the BHP for the producers varies between 5 MPa and 25 MPa. The total production period is 600 days, divided into six intervals of 100 days each. This setup results in 30 decision variables. Table 1 provides the economic parameters used to calculate the net present value (NPV).

Table 1. The values of parameters for calculating NPV.

Parameters	2	Value
Water disposal	r_w	USD 3/STB
Water injection	c_{wi}	USD 2/STB
Oil revenue	r_o	USD 60/STB
Annual discount rate	d	0.05

The accuracy and computational burden of the surrogate ensemble depends significantly on the size of the selected models [35]. To determine the optimal ensemble size, we examined the RMSE [36] and R^2 [37] indexes for ensemble sizes up to 100. Both NPV and COP were tested to identify the appropriate ensemble size and validate the performance of

the proposed selective surrogate ensemble strategy. For this analysis, 200-well control samples were generated using Latin hypercube sampling, and the objectives were evaluated by a simulator to construct the dataset. Subsequently, 100 subsets of 150 samples each were generated via bootstrap sampling, and 100 SVR models were built using Algorithm 1. Models were randomly selected from the set to form surrogate ensembles, and their accuracy was blind tested using 50 samples.

As shown in Figure 4, RMSE decreased as the ensemble size increased, stabilizing when the size exceeded 40, while R^2 continued to improve, achieving significant precision around an ensemble size of 4. However, the computational burden increased linearly with ensemble size. Based on these observations, an ensemble size of 40 was chosen, providing a balance between sufficient accuracy and manageable computational load.

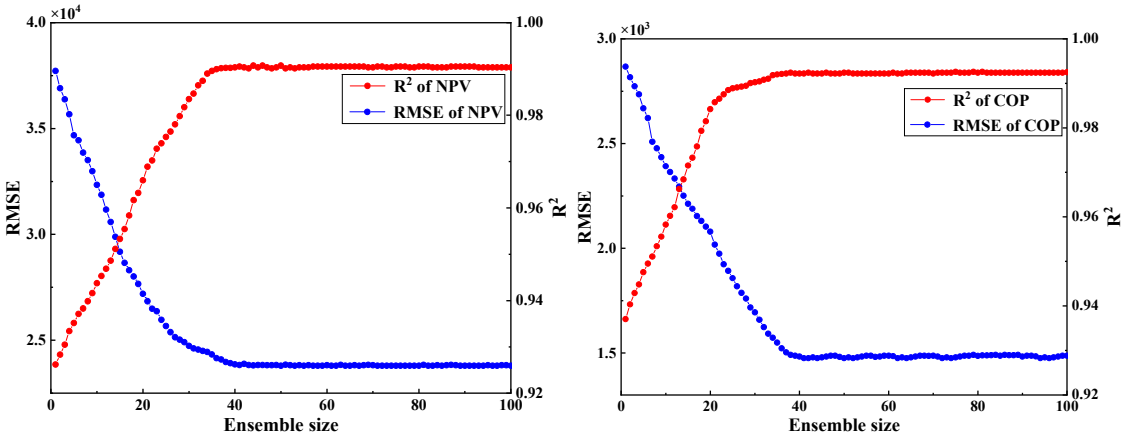


Figure 4. Changes in RMSE and R^2 of NPV (**left**) and COP (**right**) with ensemble size.

Based on the dataset, bootstrap sampling was performed for 50 subsets with 150 samples, and 50 SVR models were built to establish a surrogate ensemble for optimization. The specific parameter settings of the optimization process are shown in Table 2.

Table 2. Values of parameters used in NSGA-II.

Parameters	Value
Number of individuals	50
Iteration steps	100
Probability of mutative	0.005
Probability of cross	0.65

Figure 5 shows the Pareto fronts of the SVR-NSGA-II and the MOO-SESA. The gray hollow circles display the trajectories of intermediate solutions of the MOO-SESA. The red solid circles represent the PF of the MOO-SESA. The blue solid circles are the PF of the SVR-NSGA-II. For the same number of iterations, the MOO-SESA framework produced a PF with superior diversity and convergence compared to the SVR-NSGA-II method. Furthermore, MOO-SESA resulted in an increased Pareto front compared to that achieved using SVR-NSGA-II. To that end, the optimization results of MOO-SESA were better than SVR-NSGA-II. It is worth mentioning that the major calculation burden of MOO-SESA in this example is only 200 times the numerical simulation run plus the very small amount of time required to train 50 SVR models. However, without the surrogate model, the computational burden runs 5000 simulations in this case, which is nearly 25 times slower than the MOO-SESA method. Meanwhile, the cost of the surrogate modeling and selection process is nearly 5% of the runtime of the numerical simulation model. For a decision maker, any solutions of well control schemes in the PF could be a perfect result.

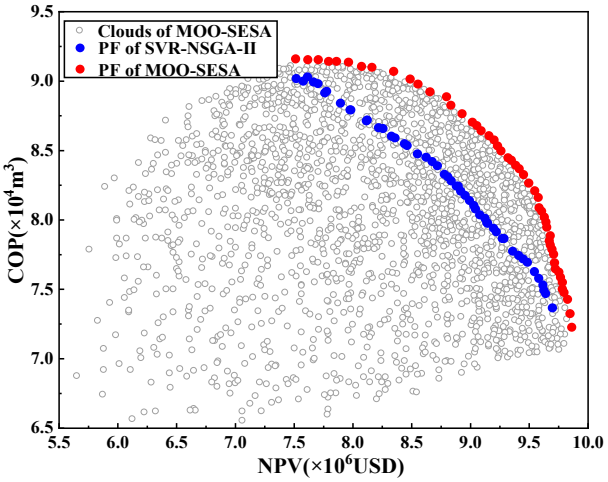


Figure 5. Pareto fronts of SVR-NSGA-II and MOO-SESA in case one.

For the purpose of further assessing the effectiveness of the MOO-SESA method, two optimization schemes—maximizing NPV (case 1) and maximizing COP (case 2)—were selected for detailed analysis. The detailed schemes for both cases are illustrated in Figure 6, while their corresponding NPV and COP values are shown in Figure 7. Case 1 achieved an NPV of USD 9.69 × 10^6 with a COP of 7.36 × 10^4 m^3, whereas case 2 resulted in an NPV of USD 7.52 × 10^6 and a COP of 9.02 × 10^4 m^3. This inverse relationship between NPV and COP clearly demonstrates the key characteristic of multi-objective optimization, where improving one objective often leads to a reduction in another.

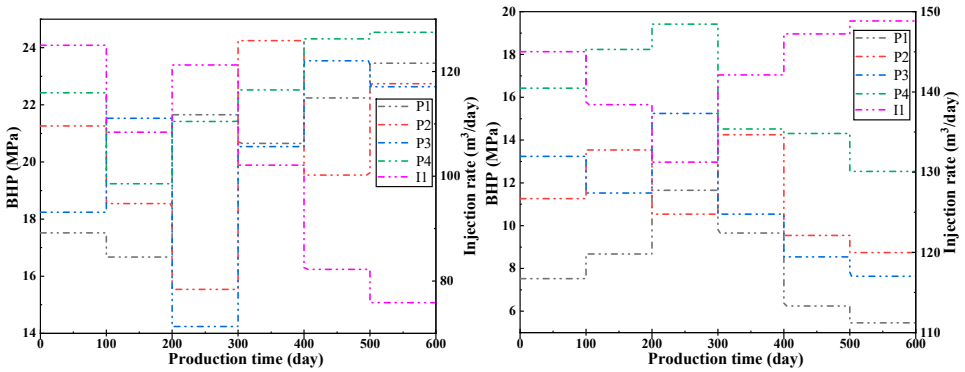


Figure 6. Detailed scheme of case 1 (**left**) and case 2 (**right**).

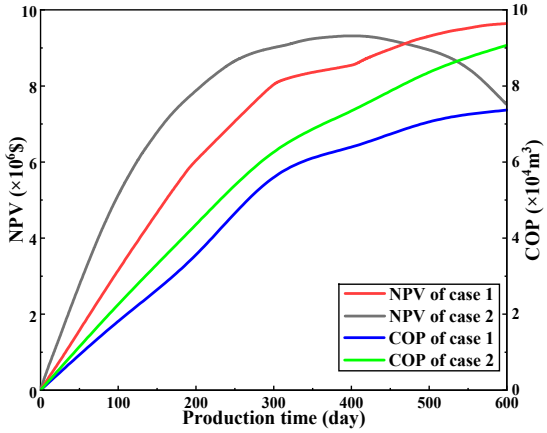


Figure 7. Curves of NPV and COP of cases.

4.2. Case 2

In this case, Egg model, which is a small synthetic reservoir model proposed by Jansen et al. (2014) [33], is selected as the 3D model to validate the quality and efficiency of the proposed MOO-SESA method. The grid block of this model is 60 × 60 × 7, totaling 25,200 grids, of which the active cells are 18,553. Because there are no active grids at the edge of the model, the model appears as an egg. For detailed and more specific parameters of the Egg model, one can refer to the work published by Jansen et al. (2014) [33]. The detailed permeability field is shown in Figure 8.

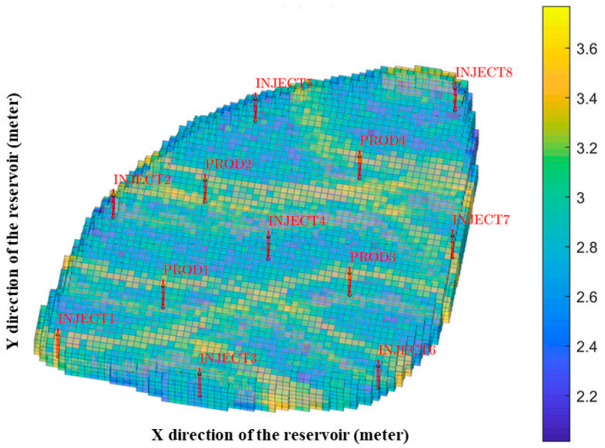


Figure 8. The detailed permeability field and well location of case 2.

In this scenario, the production wells were configured to operate at a constant pressure of 38 MPa. Consequently, the decision variables are the injection rates of the eight injectors, each bounded between 0 m^3/day and 70 m^3/day. The total production period spanned 3600 days, divided into twenty intervals of 180 days each. This setup resulted in 160 decision variables. The economic parameters for calculating the net present value (NPV) were the same as those used in the first case.

To determine the optimal ensemble scale, RMSE and R^2 metrics were used to evaluate the accuracy of ensembles of up to 100 members. The study focused on NVP and COP, with decision variables totaling 160, significantly more than in the 2D case. Using Latin hypercube sampling (LHS), 1000-well control samples were generated and evaluated by the simulator to create the dataset. Subsequently, 100 SVR models were built from bootstrap samples of 950, and another 50 samples were used for blind testing. The RMSE and R^2 trends with varying ensemble sizes are depicted in Figure 9. As shown in Figure 9, RMSE decreased as the ensemble size increased, but stabilized within a narrow range once the ensemble size exceeded 48. Similarly, R^2 improved with larger ensemble sizes, achieving considerable precision around an ensemble size of 35. However, the linear increase in computational complexity with larger ensemble sizes remained a challenge. Based on these findings, an ensemble size of 48 was selected to balance sufficient accuracy with a relatively modest computational burden.

Using bootstrap sampling, 60 subsets with 950 samples were generated to build 60 SVR models, forming the surrogate ensemble for optimization. The optimizer's parameters were consistent with those in the first case. As illustrated in Figure 10, the Pareto fronts obtained by the MOO-SESA method outperformed those of the SVR-NSGA-II, providing better diversity and convergence within the same number of iterations. The MOO-SESA method also significantly reduced the computation time, with ensemble surrogate models running in under 10 s, compared to over 300 s for the numerical simulation on an Intel Core i7-7800 CPU. This approach makes high-dimensional, computationally expensive optimization feasible, offering decision-makers a broader range of solutions tailored to specific objectives, such as maximizing economic benefits or recovery ratios.

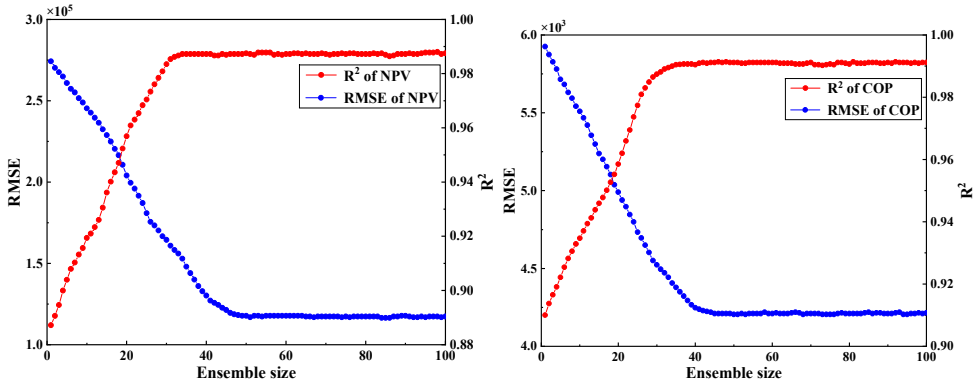


Figure 9. Change in RMSE and R^2 of NPV (**left**) and COP (**right**) with ensemble size.

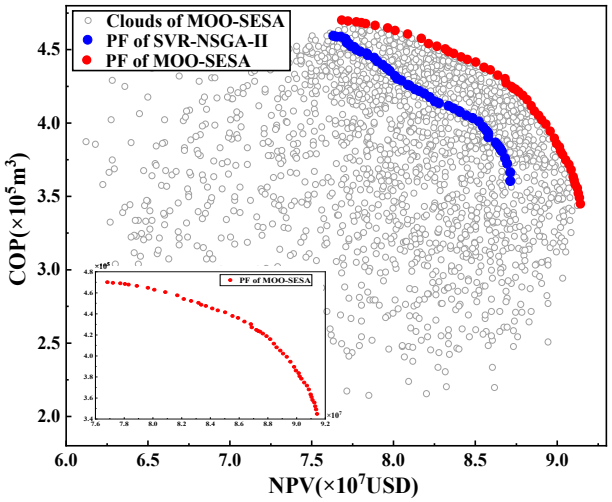


Figure 10. Pareto fronts of SVR-NSGA-II and MOO-SESA in case two.

To further assess the performance of the MOO-SESA method, we selected two extreme scenarios—one maximizing NPV (case 1) and the other maximizing COP (case 2)—to illustrate the details of the optimized solutions. The well control schemes for cases 1 and 2 during each production interval are depicted in Figure 11.

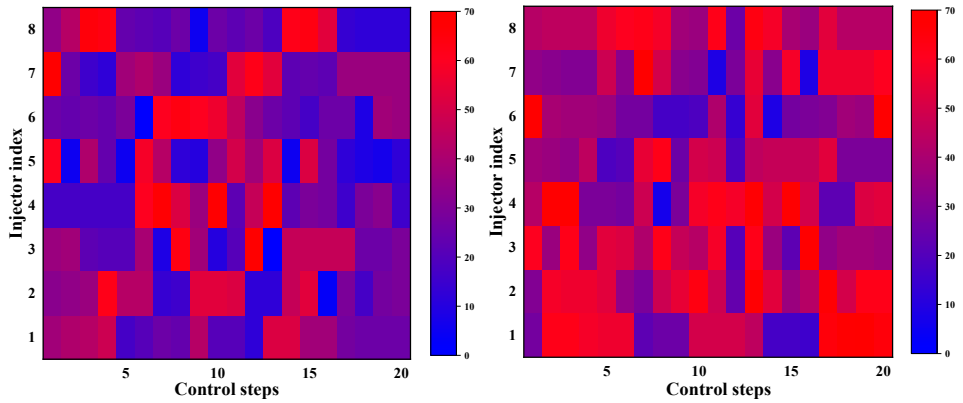


Figure 11. Well control scheme of case 1 (**left**) and case 2 (**right**).

The specific NPV and COP curves are displayed in Figure 12. Cases 1 and 2 yielded NPVs of USD 9.14×10^7 and USD 7.68×10^7, respectively. Meanwhile, for COP, cases 1 and 2 produced 3.45×10^5 m^3 and 4.71×10^5 m^3, respectively. In essence, the maximum NPV corresponded to the minimum COP, and vice versa, illustrating a fundamental characteristic

of multi-objective optimization (MOO). This means that improving one objective inevitably leads to a decline in one or more other objectives.

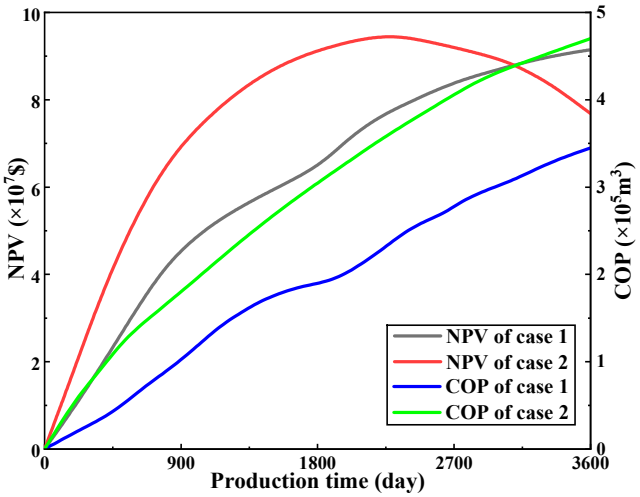


Figure 12. Specific NPV and COP curves of case 1 and 2.

5. Conclusions

This study introduced the MOO-SESA framework, an innovative optimization method designed to address the computational challenges of well control optimization problems. Our comparative analysis, which was applied to both the SPE10 2D and Egg 3D reservoir models, demonstrated that MOO-SESA significantly outperformed SVR-NSGA-II across multiple metrics. Specifically, the MOO-SESA framework delivered more efficient and superior well control scenarios with enhanced diversity and convergence of the Pareto front. This led to faster convergence rates, greater accuracy in capturing the trade-offs between competing objectives, and ultimately better decision-making for well control optimization. The results confirmed that the MOO-SESA framework's adaptability and superior performance are not limited to specific reservoir types or dimensions; it proved effective in both 2D and 3D models. This makes the MOO-SESA framework highly versatile and applicable across a range of reservoir simulation challenges, from simple, low-dimensional problems to more complex, high-dimensional ones. By significantly reducing computational costs while improving optimization quality, MOO-SESA holds tremendous promise for more efficient and robust decision-making in petroleum reservoir management. Moreover, the MOO-SESA framework holds potential for broader applications within petroleum engineering, including well placement and fracturing optimization. Furthermore, integrating MOO-SESA with real-time optimization frameworks is also worth investigating. Future research will focus on developing more accurate and efficient surrogate models and optimizing algorithms that are better suited to approaching high-dimensional, computationally complex multi-objective reservoir optimization challenges.

Author Contributions: Conceptualization, L.W. and R.D; methodology, L.W. and L.Z. (Liehui Zhang); software, L.W., R.D. and L.Z.; validation, J.Q. and C.D.A.; formal analysis, X.Z.; investigation, L.W. and B.X.; resources, H.W. and L.Z. (Liang Zhang); data curation, R.D. and L.Z. (Liang Zhang); writing—original draft preparation, L.W.; writing—review and editing, X.L., C.D.A. and R.D.; visualization, L.Z. (Liehui Zhang); supervision, H.W. and L.Z. (Liehui Zhang); project administration, J.Q.; funding acquisition, J.Q., H.W. and L.Z. (Liang Zhang). All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by [Evaluation of Porous Carbonate Reservoirs and Fundamental Research on Waterflooding Simulation] grant number [J2301], [Integrated Data-Driven and Spatio-Temporal Simulation Inversion for Complex Carbonate Reservoir Production Development Strategies] grant number [R2302] and [Basic research on evaluation of porous carbonate reservoirs and simulation of reservoir water flooding] projects.

Data Availability Statement: All data supporting the findings of this study have been included in this published article.

Acknowledgments: This research was supported by Zhenhua Oil Co., Ltd.

Conflicts of Interest: Authors Lian Wang, Rui Deng, Liang Zhang, Jianhua Qu, Hehua Wang, Xing Zhaoand and Bing Xu were employ by Zhenhua Oil Co., Ltd. Author Xindong Lv was employed by Chengdu North Petroleum Exploration and Development Technology Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Nomenclature

u	The decision variable vector.
r_o	The oil price (USD/STB).
r_w	The cost of disposing produced water (USD/STB).
c_{wi}	The water-injection cost (USD/STB).
N_t	The production steps
t_n	The n-th step
b	The annual discount rate
$\triangle t_n$	The length of the n-th step (days)
P	The number of production wells
I	The number of injection wells
$q_{o,j}^n$	The oil rates at the j-th well at the n-th step (m^3/day).
$q_{w,j}^n$	The water rates at the j-th well at the n-th step (m^3/day).
$q_{wi,j}^n$	The water injection rates at the j-th well at the n-th step (m^3/day).
MOO-SESA	Surrogate-assisted multi-objective optimization framework
NPV	Net present value
COP	Cumulative oil production

References

1. Ding, S.; Jiang, H.; Liu, G.; Sun, L.; Lu, X.; Zhao, L. Determining the levels and parameters of thief zone based on automatic history matching and fuzzy method. *J. Pet. Sci. Eng.* **2016**, *138*, 138–152. [CrossRef]
2. Wang, L.; Yao, Y.; Luo, X.; Adenutsi, C.D.; Zhao, G.; Lai, F. A critical review on intelligent optimization algorithms and surrogate models for conventional and unconventional reservoir production optimization. *Fuel* **2023**, *350*, 128826. [CrossRef]
3. Liu, X.; Reynolds, A.C. Robust gradient-based multiobjective optimization for the generation of well controls to maximize the net-present-value of production under geological uncertainty. *Math. Geosci.* **2017**, *49*, 361–394. [CrossRef]
4. Zhang, L.; Zhang, T.; Yulong, Z.; Hu, H.; Wen, S.; Wu, J.; Cao, C.; Wang, Y.; Fan, Y. A review of interaction mechanisms and microscopic simulation methods for CO_2-water-rock system. *Pet. Explor. Dev.* **2024**, *51*, 223–238. [CrossRef]
5. Wang, L.; Yao, Y.; Zhao, G.; Adenutsi, C.D.; Wang, W.; Lai, F. A hybrid surrogate-assisted integrated optimization of horizontal well spacing and hydraulic fracture stage placement in naturally fractured shale gas reservoir. *J. Pet. Sci. Eng.* **2022**, *216*, 110842. [CrossRef]
6. Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T.A.M.T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **2002**, *6*, 182–197. [CrossRef]
7. Zhang, T.; Luo, S.; Zhou, H.; Hu, H.; Zhang, L.; Zhao, Y.; Li, J.; Javadpour, F. Pore-scale modelling of water sorption in nanopore systems of shale. *Int. J. Coal Geol.* **2023**, *273*, 104266. [CrossRef]
8. Fonseca, R.M.; Reynolds, A.C.; Jansen, J.D. Generation of a Pareto front for a biobjective water flooding optimization problem using approximate ensemble gradients. *J. Petrol. Sci. Eng.* **2016**, *147*, 249–260. [CrossRef]
9. Feng, D.; Bakhshian, S.; Wu, K.; Song, Z.; Ren, B.; Li, J.; Hosseini, S.A.; Li, X. Wettability effects on phase behavior and interfacial tension in shale nanopores. *Fuel* **2021**, *290*, 119983. [CrossRef]
10. Liu, X.; Reynolds, A.C. Gradient-based multiobjective optimization for maximizing expectation and minimizing uncertainty or risk with application to optimal well-control problem with only bound constraints. *SPE J.* **2016**, *21*, 1813–1829. [CrossRef]
11. Fonseca, R.; Leeuwenburgh, O.; Rossa, E.D.; Van den Hof, P.M.; Jansen, J.D.D. Ensemble-Based Multiobjective Optimization of on/off Control Devices under Geological Uncertainty. *SPE Res. Eval. Eng.* **2015**, *18*, 554–563. [CrossRef]
12. Feng, D.; Li, X.; Wang, X.; Li, J.; Zhang, X. Capillary filling under nanoconfinement: The relationship between effective viscosity and water-wall interactions. *Int. J. Heat Mass Transf.* **2018**, *118*, 900–910. [CrossRef]
13. Liu, Z.; Reynolds, A.C. An SQP-Filter Algorithm with an Improved Stochastic Gradient for Robust Life-Cycle Optimization Problems with Nonlinear Constraints. In Proceedings of the Paper Presented at the SPE Reservoir Simulation Conference, Galveston, TX, USA, 10–11 April 2019. SPE-193925-MS.

14. Ishibuchi, H.; Shibata, Y. A Similarity-Based Mating Scheme for Evolutionary Multi-Objective Optimization. In Proceedings of the Genetic and Evolutionary Computation Conference, Chicago, IL, USA, 12–16 July 2003; pp. 1065–1076.
15. Chang, Y.; Bouzarkouna, Z.; Devegowda, D. Multi-objective optimization for rapid and robust optimal oilfield development under geological uncertainty. *J. Comput. Geosci.* **2015**, *19*, 933–950. [CrossRef]
16. Yasari, E.; Pishvaie, M.R.; Khorasheh, F.; Salahshoor, K.; Kharrat, R. Application of multi-criterion robust optimization in water-flooding of oil reservoir. *J. Petrol. Sci. Eng.* **2013**, *109*, 1–11. [CrossRef]
17. Luo, J.; Qi, Y.; Xie, J.; Zhang, X. A hybrid multi-objective PSO–EDA algorithm for reservoir flood control operation. *Appl. Soft Comput.* **2015**, *34*, 526–538. [CrossRef]
18. Fu, J.; Wen, X.H. Model-based multiobjective optimization methods for efficient management of subsurface flow. *SPE J.* **2017**, *22*, 1984–1998. [CrossRef]
19. Jansen, J.D.; Durlofsky, L.J. Use of Reduced-Order Models in Well Control Optimization. *Optimiz. Eng.* **2017**, *18*, 105–132. [CrossRef]
20. Zhao, H.; Kang, Z.; Zhang, X.; Sun, H.; Cao, L.; Reynolds, A.C. A Physics-Based Data-Driven Numerical Model for Reservoir History Matching and Prediction with a Field Application (Associated Discussion Available as Supporting Information). *SPE J.* **2016**, *21*, 2175–2194. [CrossRef]
21. Cao, F.; Luo, H.; Lake, L.W. Oil-Rate Forecast by Inferring Fractional-Flow Models from Field Data with Koval Method Combined with the Capacitance/Resistance Model. *SPE Res. Eval. Eng.* **2015**, *18*, 534–553. [CrossRef]
22. Yousef, A.A.; Gentil, P.H.; Jensen, J.L.; Lake, L.W. A Capacitance Model to Infer Interwell Connectivity from Production and Injection Rate Fluctuations. *SPE J.* **2006**, *9*, 630–646. [CrossRef]
23. Guo, Z.; Reynolds, A.C. Robust Life-Cycle Production Optimization with a Support-Vector-Regression Proxy. *SPE J.* **2018**, *23*, 2409–2427. [CrossRef]
24. Guo, Z.; Reynolds, A.C.; Zhao, H. A Physics-Based Data-Driven Model for History Matching, Prediction, and Characterization of Water-flooding Performance. *SPE J.* **2018**, *23*, 367–395. [CrossRef]
25. Zhang, L.; Li, Z.; Lai, F.; Li, H.; Adenutsi, C.D.; Wang, K.; Xu, W. Integrated optimization design for horizontal well placement and fracturing in tight oil reservoirs. *J. Pet. Sci. Eng.* **2019**, *178*, 82–96. [CrossRef]
26. Ashena, R.; Thonhauser, G. Application of artificial neural networks in geoscience and petroleum industry. In *Artificial Intelligent Approaches in Petroleum Geosciences*; Springer International Publishing: Cham, Switzerland, 2024; pp. 115–154.
27. Chu, M.G.; Min, B.; Kwon, S.; Park, G.; Kim, S.; Huy, N.X. Determination of an infill well placement using a data-driven multi-modal convolutional neural network. *J. Pet. Sci. Eng.* **2020**, *195*, 106805. [CrossRef]
28. Wang, L.; Li, Z.P.; Adenutsi, C.D.; Zhang, L.; Lai, F.; Wang, K. A novel multi-objective optimization method for well control parameters based on PSO-LSSVR proxy model and NSGA-II algorithm. *J. Pet. Sci. Eng.* **2020**, *196*, 107694. [CrossRef]
29. Wang, L.; Yao, Y.; Wang, K.; Adenutsi, C.D.; Zhao, G.; Lai, F. A novel surrogate-assisted multi-objective optimization method for well control parameters based on tri-training. *Nat. Resour. Res.* **2021**, *30*, 4825–4841. [CrossRef]
30. Mendes-Moreira, J.; Soares, C.; Jorge, A.M.; Sousa, J.F.D. Ensemble Approaches for Regression: A Survey. *ACM Comput. Surv.* **2013**, *45*, 10. [CrossRef]
31. Suen, Y.L.; Melville, P.; Mooney, R.J. Combining bias and variance reduction techniques for regression trees. In Proceedings of the European Conference on Machine Learning, Porto, Portugal, 3–7 October 2005.
32. Zhou, Z.H. *Ensemble Methods: Foundations and Algorithms*; Taylor & Francis: Abingdon, UK, 2012.
33. Jansen, J.D.; Fonseca, R.M.; Kahrobaei, S.; Siraj, M.M.; Van Essen, G.M.; Van den Hof, P.M.J. The Egg Model-A Geological Ensemble for Reservoir Simulation. *Geosci. Data J.* **2014**, *1*, 192–195. [CrossRef]
34. Lie, K.-A. *An Introduction to Reservoir Simulation Using MATLAB/GNU Octave: User Guide for the MATLAB Reservoir Simulation Toolbox (MRST)*; Cambridge University Press: Cambridge, UK, 2019.
35. Crone, S.F.; Guajardol, J.; Weber, R. The Impact of Preprocessing on Support Vector Regression and Neural Networks in Time Series Prediction. In Proceedings of the International Conference on Data Mining DMIN '06, Las Vegas, NV, USA, 26–29 June 2006; pp. 37–44.
36. Drucker, H.; Burges, C.J.C.; Kaufman LSmola, A.; Vapnik, V. Support Vector Regression Machines. In *Neural Information Processing Systems*; Mozer, M.C., Jordan, J.I., Petsche, T., Eds.; MIT Press: Cambridge, MA, USA, 1997; Volume 9, pp. 155–161.
37. Guo, Z.; Reynolds, A.C.; Zhao, H. Waterflooding Optimization with the INSIM-FT Data-Driven Model. *Computat. Geosci.* **2018**, *22*, 745–761. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.