*energies*

MDPI

*Article*

# Prediction of Oil Recovery Factor in Stratified Reservoirs after Immiscible Water-Alternating Gas Injection Based on PSO-, GSA-, GWO-, and GA-LSSVM

**Pål Østebø Andersen [1,*], Jan Inge Nygård [2] and Aizhan Kengessova [1,3]**

[1] Department of Energy Resources, Faculty of Science and Technology, University of Stavanger, 4021 Stavanger, Norway; aizhankeness@gmail.com
[2] Bouvet, 4020 Stavanger, Norway; jan.nygard@bouvet.no
[3] Timal Consulting Group LLP, Atyrau 060011, Kazakhstan
* Correspondence: pal.andersen@uis.no

**Abstract:** In this study, we solve the challenge of predicting oil recovery factor (RF) in layered heterogeneous reservoirs after 1.5 pore volumes of water-, gas- or water-alternating-gas (WAG) injection. A dataset of ~2500 reservoir simulations is analyzed based on a Black Oil 2D Model with different combinations of reservoir heterogeneity, WAG hysteresis, gravity influence, mobility ratios and WAG ratios. In the first model MOD1, RF is correlated with one input (an effective WAG mobility ratio $M^*$). Good correlation (Pearson coefficient −0.94), but with scatter, motivated a second model MOD2 using eight input parameters: water–oil and gas–oil mobility ratios, water–oil and gas–oil gravity numbers, a reservoir heterogeneity factor, two hysteresis parameters and water fraction. The two mobility ratios exhibited the strongest correlation with RF (Pearson coefficient −0.57 for gas-oil and −0.48 for water-oil). LSSVM was applied in MOD2 and trained using different optimizers: PSO, GA, GWO and GSA. A physics-based adaptation of the dataset was proposed to properly handle the single-phase injection. A total of 70% of the data was used for training, 15% for validation and 15% for testing. GWO and PSO optimized the model equally well ($R^2$ = 0.9965 on the validation set), slightly better than GA and GSA ($R^2$ = 0.9963). The performance metrics for MOD1 in the total dataset were: RMSE = 0.050 and $R^2$ = 0.889; MOD2: RMSE = 0.0080 and $R^2$ = 0.998. WAG outperformed single-phase injection, in some cases with 0.3 units higher RF. The benefits of WAG increased with stronger hysteresis. The LSSVM model could be trained to be less dependent on hysteresis and the non-injected phase during single-phase injection.

**Keywords:** water-alternating-gas (WAG); physics-informed machine learning; least square support vector machine (LSSVM); particle swarm optimization (PSO); dimensionless numbers; hysteresis; genetic algorithm (GA); gravitational search algorithm (GSA); grey wolf optimization (GWO)



**Citation:** Andersen, P.Ø.; Nygård, J.I.; Kengessova, A. Prediction of Oil Recovery Factor in Stratified Reservoirs after Immiscible Water-Alternating Gas Injection Based on PSO-, GSA-, GWO-, and GA-LSSVM. *Energies* **2022**, *15*, 656. https://doi.org/10.3390/en15020656

Academic Editor: Dameng Liu

Received: 5 October 2021
Accepted: 11 January 2022
Published: 17 January 2022



## 1. Introduction

Oil recovery through water or gas injection often lacks efficiency due to the unfavorable mobility ratio between the oil and the displacing phase. Viscous fingering, gravity segregation and heterogeneity can also lead to poor sweep. Gas features low viscosity and density and can achieve channeling and early breakthrough [1,2]. Water-alternating gas injection (WAG) is an enhanced oil recovery (EOR) technique in which water and gas are injected in cycles to displace the oil. This type of technique mitigates the exponential rate decline seen in most fields after peak production [3]. The mobility of each injected fluid is reduced by the presence of the other, producing a more favorable mobility ratio to oil. Gravity segregation becomes less detrimental, since gas sweeps the top of the reservoir, while water sweeps the bottom [2,4]. Gas usually results in lower residual oil saturation than water, but this can be further lowered by WAG. Thus, WAG utilizes the advantages of water and gas injection and minimizes their individual downsides. In several field

implementations, it has been beneficial to use WAG; the oil recovery factor in 59 fields increased when WAG was introduced, by 5% to 10 % of the oil originally in place [2]. Sanchez [5] reported that 80% of US WAG projects were beneficial. Micromodel studies also demonstrate better oil recovery with WAG than with single-phase displacement [6].

WAG introduces more design and operational parameters compared to water or gas injection, such as the WAG ratio (volume water to volume gas injected), number of cycles, cycle volume, injection rates and pressures. This may further affect optimal well placements. A 1:1 WAG ratio is considered common or even optimal [4]. Variation in WAG ratio with project time (tapering) has been conducted, partly due to limited gas access, for limiting gas production or for optimization. [7] used ensemble-based optimization of injector and producer well controls at each WAG cycle to maximize the net present value for a channeled reservoir model. Whether the reservoir pressure is above the system's minimum miscibility pressure (MMP) determines whether the gas is miscible or immiscible with the oil [8]. During miscible displacement, the oil and gas become practically the same phase and residual oil saturation can approach zero. Kulkarni [9] found that miscible gas ($CO_2$) core flooding (continuous or WAG) outperformed immiscible injection. The choice or modification of the injected fluids can also improve the outcome. Foam, surfactant, polymer, low salinity brine and $CO_2$ are some alternatives [4].

Reservoir geology or heterogeneity is important during any field development. During gas injection, heterogeneity in terms of thief zones, stratification or fractures can cause gas channeling and early breakthrough, which can be mitigated by WAG [4]. Favorable well placement relative to the dip angle can provide more stable frontal displacement of oil. In heterogeneous reservoirs, gravity and capillary forces can divert flow from highly permeable layers to less permeable layers. These effects are more important in naturally fractured reservoirs, where advective forces are unable to mobilize oil [10,11].

The simultaneous flow of oil, water and gas requires the detailed measurement, quantification and correlation of three-phase relative permeabilities [12–14]. Injecting water and gas alternately causes gas and water saturations to rise and fall, resulting in hysteresis [15–19]. During WAG, the relative permeability of gas is more affected by hysteresis than oil and water [4] and hysteresis tends to decrease gas mobility. This reduction delays gas breakthrough and reduces gravity segregation. The Land [12] and Carlson [15] models are widely used to model relative permeability hysteresis.

Machine learning (ML) has gained increased popularity in the petroleum industry in recent years. ML algorithms can be useful for understanding trends in complex datasets and provide multivariate nonlinear regression or classification. Their applications include lithology classification [20], selecting EOR methods [21], locating optimal drilling spots [22], correlating asphaltene precipitation [23] or predicting $CO_2$ viscosity [24]. Important steps in developing ML models include selecting appropriate input and output variables, acquiring sufficient quality data, applying a suitable ML algorithm and tuning its metaparameters to prevent over- or under-fitting, usually via optimization algorithms.

This study makes use of the least squares support vector machine (LSSVM) algorithm, based on the works [25–27], for nonlinear regression. This algorithm has been applied in many contexts, such as predicting drilling fluid density [28], gas solubility [29–32], water availability [33], energy consumption [34,35], shale gas adsorption [36], wind power [37,38] and even tourism flow [39]. LSSVM has been successfully combined with optimizers such as particle swarm optimization (PSO), genetic algorithm (GA) and grey wolf Optimization (GWO) and has, in many cases, outperformed regression algorithms such as artificial neural networks, radial basis function, gene expression programming and adaptive neuro-fuzzy interference system [29,30,40,41].

In recent studies, [42,43] simulated $CO_2$ WAG injection in a reservoir model and developed machine learning proxy models with different algorithms to predict current rates of oil, gas and water based on current time, gas and water injection rates, half cycle time and operational constraints. The recovery factor and cumulative production were calculated from the produced output. The calibrated proxy models were used to optimize

the WAG process. [41] used LSSVM and other ML approaches to predict two-phase relative permeabilities and combined them via correlations in previous research to estimate three-phase relative permeabilities and the performance of a WAG core flood. [40] correlated the oil recovery performance of EOR carbonated water injection using LSSVM. [44] used ML to optimize well placement during WAG injection. [45,46] used ML to co-optimize $CO_2$ injection for oil recovery and storage during WAG under different operational constraints.

In this study, our main contribution is to predict the reservoir oil recovery factor (RF) in layered reservoirs during immiscible WAG and single-phase (gas or water) injection for different fluid, reservoir, geometrical and operational conditions. This is a relatively complex task given the number of parameters involved and their coupled nature. Based on a comprehensive simulation database generated in both [47] and this work, we present two predictive approaches. In the first (MOD1), a dimensionless number $M^*$, derived from Nygård and Andersen's study [47], is applied as a single input parameter. The second approach (MOD2) applies eight physics-motivated dimensionless input parameters to improve predictive power compared to the first method: two mobility ratios, two gravity numbers, injected water fraction, reservoir heterogeneity factor and two hysteresis parameters. In both models, the input variables incorporate all the system information. The latter approach, MOD2, utilizes the ML regression algorithm, LSSVM, with metaparameters optimized by either PSO (Particle Swarm Optimization), GSA (Gravity Search Algorithm), GA (Genetic Algorithm) or GWO (Grey Wolf Optimization). LSSVM has been optimized successfully in other works using PSO [20,23,36], GSA [37,38], GA [28] and GWO [29,32,33]. We propose a methodology to ensure the physical behavior of the machine learning model MOD2. We then adapt the dataset to be independent of hysteresis and the non-injected phase when single-phase injection is performed. Some of the research questions we investigate are:

- How well do the models predict WAG performance?
- Which parameters affect RF the most?
- Do the parameters have a positive or negative effect on RF?
- Will the models properly account for WAG injection and single phase injection?

The paper is structured as follows. The model serving as the basis for the simulation results is outlined in Section 2.1. The dimensionless number $M^*$ is outlined in Section 2.2. This number is used in the single-input parameter model. The machine learning approach and dataset follow in Section 2.3. The eight input parameters of the second approach are also presented in those latter sections. The results from analyzing the data are shown in Section 3 and the paper is concluded in Section 4.

## 2. Theory

### *2.1. Mathematical Model*

We consider the same modeling approach for immiscible WAG injection as [47]: A 2D reservoir layered in a horizontal direction, with one injector and one producer, both vertical and perforated along the full reservoir height. See Figure 1 for an illustration. A black oil model is assumed with an incompressible and immiscible three-phase flow of oil, water and gas and negligible capillary pressure. WAG is applied from the start, rather than as a tertiary method. Relevant equations are presented below:

$$\partial_t(\phi s_o) = \partial_x(f_o u_{Tx}) + \partial_z(f_o u_{Tz}) + \partial_z(K_z g \lambda_w f_o \Delta\rho_{wo}) - \partial_z\left(K_z g \lambda_g f_o \Delta\rho_{og}\right) \quad (1)$$

$$\partial_t(\phi s_w) = \partial_x(f_w u_{Tx}) + \partial_z(f_w u_{Tz}) - \partial_z(K_z g \lambda_o f_w \Delta\rho_{wo}) - \partial_z\left(K_z g \lambda_g f_w \Delta\rho_{wg}\right) \quad (2)$$

$$\partial_x u_{Tx} + \partial_z u_{Tz} = 0, \quad (3)$$

$$u_{Tx} = -K_x \lambda_T \partial_x p,\ u_{Tz} = -K_z \lambda_T \partial_z p + K_z g\left(\lambda_o \rho_o + \lambda_w \rho_w + \lambda_g \rho_g\right) \quad (4)$$

where $\phi$ denotes porosity, $s_i$ saturation of phase $i$, $f_i$ fractional flow, $K$ permeability, $\lambda_i$ mobility, $\rho_i$ density, $u_T$ total Darcy flux and $p$ pressure. Corey correlation was applied for

relative permeabilities, while gas relative permeability hysteresis was incorporated using Land's trapping model [12] (which reduces the mobile gas saturation interval based on the parameter $C$) and Carlson's hysteresis model [15] with parameter $\alpha$ (which reduces gas relative permeability).

Nygård and Andersen [47] ran simulations systematically to investigate the role of gravity segregation, the mobility ratios between the three phases, heterogeneity, hysteresis and WAG ratio and how they affected RF after 1.5 pore volumes of fluid were injected. The simulations were scaled using a combined dimensionless mobility ratio $M^*$ stating how effectively the injected fluids displaced oil under the given conditions, summarized as follows. In the design of this number, the mechanisms were incorporated one at a time. We refer to Tables A1–A3 in Appendix A for several important simulation input parameters or model configurations that were constant in the simulations. More details can be found in the original paper. The fact that these parameters remained constant was mainly due to prioritization. However, these input parameters were incorporated in the dimensionless numbers presented in the following sections.

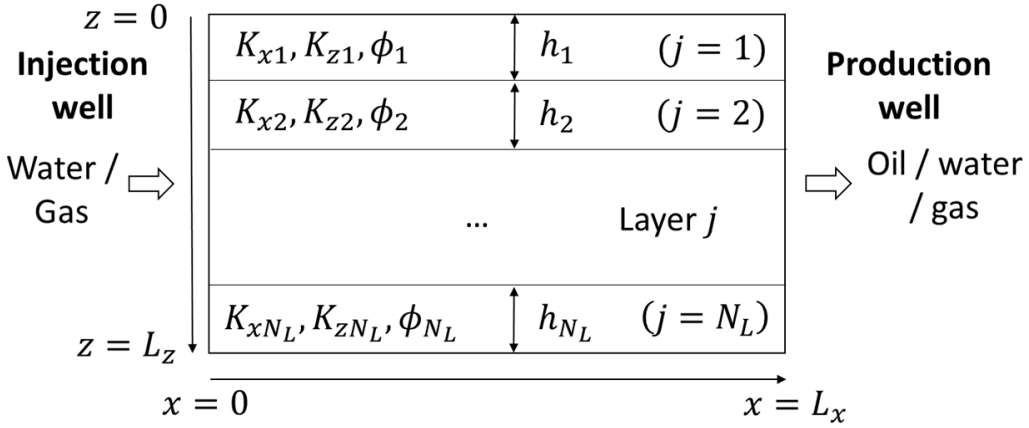


**Figure 1.** The geometrical configuration of the model (modified from [47]). is the distance from the injector, while is the distance from the top of the reservoir.

Oil recovery factor (RF) is the output parameter of interest, defined as:

$$\mathrm{RF} = \frac{volume\ oil\ produced}{volume\ oil\ initially\ in\ place} = 1 - \frac{\sum_{j=1}^{N_z}\sum_{k=1}^{N_x}\phi(z_j)s_o(x_k, z_j)}{\sum_{j=1}^{N_z}\sum_{k=1}^{N_x}\phi(z_j)s_{oi}} \tag{5}$$

Every grid block features same dimension $\Delta x \Delta z$. RF is reported after 1.5 pore volumes are injected.

### 2.2. WAG Efficiency Characterization Using Dimensionless Number

The characteristic mobility ratio $M^*$ defined by Nygård and Andersen [47] features the following functional relation:

$$M^* = \left( \frac{r_w}{M^*_{w/o} F_H F_G^{w/o}} + \frac{1 - r_w}{M^*_{g/o} F_H F_G^{g/o}} \right)^{-1} \tag{6}$$

$r_w$ is the volume fraction of water in each cycle. Larger $M^*$ is associated with lower recovery factor. Characteristic two-phase mobilities $\lambda_i^*$ for each phase $i$ were found by

averaging their mobility over their mobile saturation interval and used to define two-phase mobility ratios $M^*_{w/o}$, $M^*_{g/o}$:

$$M^*_{w/o} = \frac{\lambda^*_w}{\lambda^*_{ow}},\ \lambda^*_w = \frac{k^{max}_{rw}}{\mu_w}\frac{\left(1-\frac{s_{wr}}{s_{w,max}}\right)}{(n_w+1)},\ \lambda^*_{ow} = \frac{k^{max}_{row}}{\mu_o}\frac{\left(1-\frac{s_{orw}}{s_{ow,max}}\right)}{(n_{ow}+1)} \tag{7}$$

$$M^*_{g/o} = \frac{\lambda^*_g}{\lambda^*_{og}},\ \lambda^*_g = \frac{k^{max}_{rg}}{\mu_g}\frac{\left(1-\frac{s_{gr}}{s_{g,max}}\right)}{(n_g+1)},\ \lambda^*_{og} = \frac{k^{max}_{rog}}{\mu_o}\frac{\left(1-\frac{s_{org}}{s_{og,max}}\right)}{(n_{og}+1)} \tag{8}$$

$s_{i,max}$ denotes the saturation of phase $i$ where end-point relative permeability $k^{max}_{ri}$ is obtained, $n_i$ is the Corey exponent and $\mu_i$ is viscosity. A heterogeneity factor $F_H$ was derived from the horizontal permeability $K_{xj}$ and layer height $h_j$ distribution over layers $j = 1 : N_L$:

$$F_H = \frac{\overline{K}^{ar}_x}{\overline{K}^{ha}_x},\ \overline{K}^{ar}_x = \left(\sum_{j=1}^{N_L} h_j\right)^{-1}\sum_{j=1}^{N_L} h_j K_{x,j}, \overline{K}^{ha}_x = \left(\sum_{j=1}^{N_L} h_j\right)\left(\sum_{j=1}^{N_L}\frac{h_j}{K_{x,j}}\right)^{-1} \tag{9}$$

Two-phase gravity numbers were defined using the ratio of two-phase segregation time $t_{seg}$ and the residence time $t_{res}$ of the injected phase:

$$N^{w/o}_G = \frac{t^w_{res}}{t^{w/o}_{seg}},\ t^w_{res} = \frac{L_x L_y \sum_{j=1}^{N_L}\phi_j h_j}{Q_w},\ t^{w/o}_{seg} = \frac{H\phi}{K^{ha}_z \Delta\rho_{wo} g}\left(\frac{1}{\lambda^*_w}+\frac{1}{\lambda^*_{ow}}\right) \tag{10}$$

$$N^{g/o}_G = \frac{t^g_{res}}{t^{g/o}_{seg}},\ t^g_{res} = \frac{L_x L_y \sum_{j=1}^{N_L}\phi_j h_j}{Q_g},\ t^{g/o}_{seg} = \frac{H\phi}{K^{ha}_z \Delta\rho_{go} g}\left(\frac{1}{\lambda^*_g}+\frac{1}{\lambda^*_{og}}\right), \tag{11}$$

It was found that the role of gravity depended on heterogeneity and two-phase gravity factors $F^{w/o}_G, F^{w/o}_G$ accounting for this coupling were introduced:

$$F^{w/o}_G = \frac{1+a_1\left(N^{w/o}_G\right)^{a_2}}{1+a_1(F_H-1)\left(N^{w/o}_G\right)^{a_2}},\ F^{g/o}_G = \frac{1+a_1\left(N^{g/o}_G\right)^{a_2}}{1+a_1(F_H-1)\left(N^{g/o}_G\right)^{a_2}} \tag{12}$$

Note the unitless tuning parameters $a_1 = 3$ and $a_2 = 0.5$. Finally, hysteresis was incorporated into the gas characteristic relative permeability. Land's parameter $C$ defines a hysteresis residual gas saturation $s^{hyst}_{gr}$:

$$s^{hyst}_{gr} = s_{gr} + \frac{s_{g,max} - s_{gr}}{1 + C\left(s_{g,max} - s_{gr}\right)} \tag{13}$$

A further modification according to $r_w$ was made:

$$s^{wag}_{gr} = s_{gr}(1-r_w) + r_w s^{hyst}_{gr} \tag{14}$$

Additionally, the gas relative permeability end point $k^{max}_{rg}$ in $\lambda^*_g$, (see Equation (8)), was reduced due to hysteresis. The reductions were performed individually for the gas–oil mobility ratio and the gas–oil gravity number first based on the parameter $\alpha$ and heterogeneity factor $F_H$ using unitless tuning parameters $b_1 = 1, b_2 = 0.5$ and $b_3 = 10, b_4 = 2$:

$$k^{max,hyst}_{rg,M} = \frac{k^{max}_{rg}}{1+b_1 F^{b_2}_H \alpha},\ k^{max,hyst}_{rg,N_G} = \frac{k^{max}_{rg}}{1+b_3 F^{b_4}_H \alpha} \tag{15}$$

Next, the fraction $r_w$ was incorporated according to:

$$k_{rg,M}^{wag} = \left( \frac{1 - r_w}{k_{rg}^{max}} + \frac{r_w}{k_{rg,M}^{max,hyst}} \right)^{-1}, \; k_{rg,N_G}^{wag} = \left( \frac{1 - r_w}{k_{rg}^{max}} + \frac{r_w}{k_{rg,N}^{max,hyst}} \right)^{-1} \tag{16}$$

We then obtained the hysteresis-corrected characteristic gas mobilities $\lambda_{g,M}^*$ and $\lambda_{g,N_G}^*$ by replacing $s_{gr}$ with $s_{gr}^{wag}$ in (8), while the end-point relative permeability $k_{rg}^{max}$ in (8) was replaced by $k_{rg,M}^{wag}$ in the gas-oil mobility ratio $M_{g/o}^*$ in (8) and by $k_{rg,N_G}^{wag}$ in the gas–oil gravity number $N_G^{g/o}$ in (11):

$$\lambda_{g,M}^* = \frac{1}{\mu_g} \left( 1 - \frac{s_{gr}^{wag}}{s_{g,max}} \right) \frac{k_{rg,M}^{wag}}{n_g + 1}, \; \lambda_{g,N_G}^* = \frac{1}{\mu_g} \left( 1 - \frac{s_{gr}^{wag}}{s_{g,max}} \right) \frac{k_{rg,N_G}^{wag}}{n_g + 1} \tag{17}$$

Every input parameter is incorporated in the dimensionless number $M^*$. Note that during single-phase injection ($r_w = 0$ or $1$), the two-phase parameters involving the phase not injected do not affect $M^*$. Similarly, hysteresis does not affect $M^*$ during single-phase injection.

### *2.3. Workflow*

#### 2.3.1. Model Input Parameters

In the first model, MOD1, we take

$$x_0 = \log_{10} M^* \tag{18}$$

as the only input parameter to predict RF. We also consider a machine learning model (MOD2) in which the following eight dimensionless numbers are used as input parameters:

$$\begin{gathered} x_1 = r_w, \; x_2 = \log_{10} F_H, \; x_3 = \alpha, \; x_4 = \log_{10} C, \\ x_5 = \log_{10} M_{g/o}^*, \; x_6 = \log_{10} M_{w/o}^*, \; x_7 = \log_{10} N_G^{w/o}, \; x_8 = \log_{10} N_G^{g/o} \end{gathered} \tag{19}$$

These numbers reflect injected fluid fractions $x_1$, heterogeneity $x_2$, hysteresis $x_3, x_4$, relative magnitude of fluid mobilities $x_5, x_6$ and gravity vs. advective forces $x_7, x_8$. They incorporate all the input parameters used in the Eclipse model and the number $M^*$. The overall workflow is demonstrated in Figure 2, where the two modeling approaches after the data collection step are indicated. In MOD1, a polynomial regression is performed, while machine learning is used for MOD2. The data and the detailed steps for developing MOD2 are explained below.

#### 2.3.2. Reservoir Simulation Dataset and Model Approaches

In addition to the 1648 WAG and 96 single phase injection simulations generated by [47], 824 new WAG simulations were performed with new $C$ and $\alpha$ values combined with existing combinations of heterogeneity, density and mobility. In the previous study, $C$ and $\alpha$ were selected primarily to cover no or significant hysteresis. The values $\alpha = 0$ and $C = 1000$ were assigned to points without hysteresis influence from the respective parameters. For MOD1, each simulation allows the calculation of $M^*$, which is input to the corresponding output RF. The $1648 + 96 + 824 = 2568$ data points were analyzed with MOD1 and correlated using a polynomial expression between RF and $x_0$.

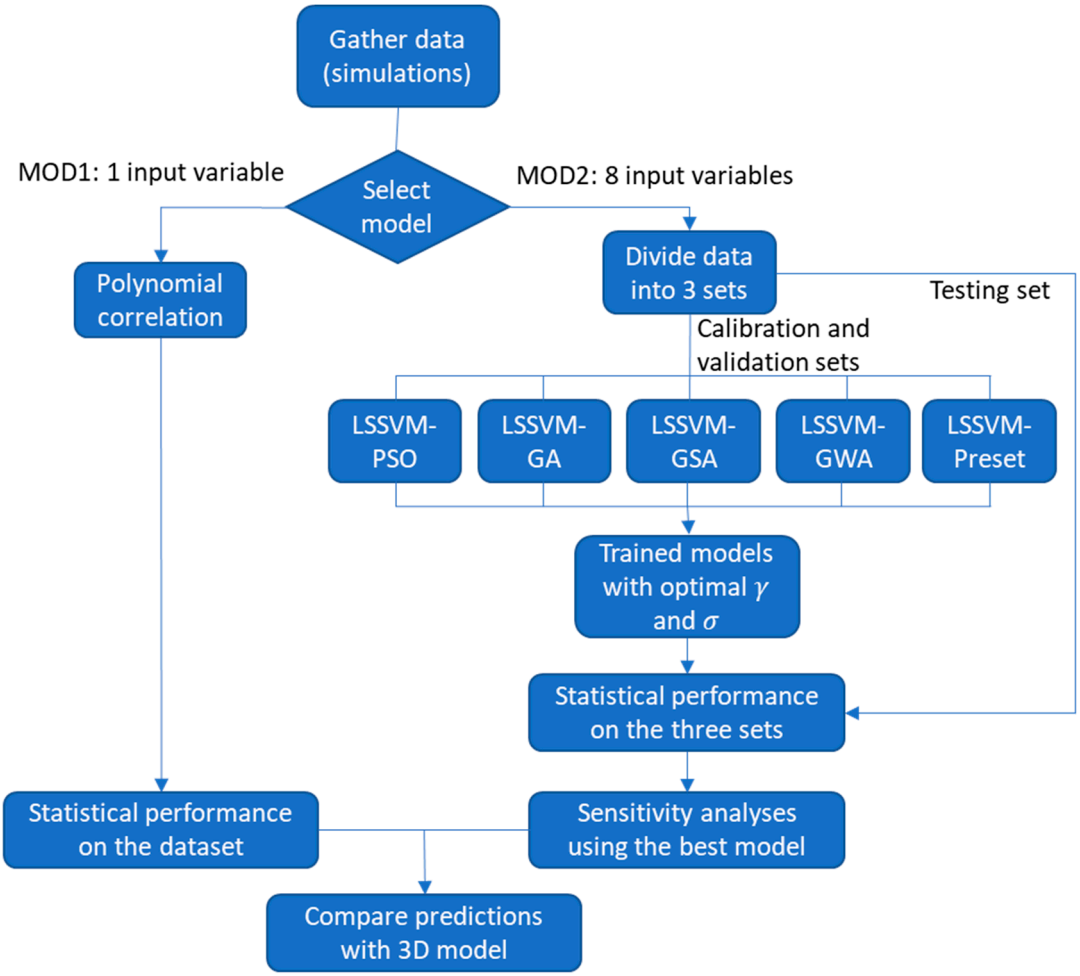


**Figure 2.** Workflow demonstrating the development, assessment and application of the models.

For the single-phase injection of water ($r_w = 1$), values for $N_G^{g/o}$ and $M_{g/o}^*$ are not well defined since gas is not injected. The case is similar regarding $N_G^{w/o}$ and $M_{w/o}^*$ at $r_w = 0$ (gas injection). Further, hysteresis parameters $C, \alpha$ should not matter. The insensitivity of $M^*$ to the mentioned parameters under single-phase injection was ensured during its derivation. To properly define the input values under single-phase injection for MOD2 the following approach was taken:

Each single-phase data point was duplicated to 16 data points, in which all combinations of high and low values of the four missing parameters were assigned. Specifically, for gas injection (points with $r_w = 0$, indexed 'g'), the following values were set for $x_3, x_4, x_6, x_7$:

$$\begin{aligned} x_{3,g} &= \overline{x}_{3,WAG} \pm X\sigma_{3,WAG},\ x_{4,g} = \overline{x}_{4,WAG} \pm X\sigma_{4,WAG}, \\ x_{6,g} &= \overline{x}_{6,WAG} \pm X\sigma_{6,WAG},\ x_{7,g} = \overline{x}_{7,WAG} \pm X\sigma_{7,WAG}, \end{aligned} \tag{20}$$

where $\overline{x}_{i,WAG}$ $(i = 3, 4, 6, 7)$ indicate the average of the data point values applied in the WAG cases and $\sigma_{i,WAG}$ $(i = 3, 4, 6, 7)$ the corresponding standard deviations. $X$ is a multiplier. Similarly, for water injection (indexed '$w$') the following values were set for $x_3, x_4, x_5, x_8$:

$$\begin{aligned} x_{3,w} &= \overline{x}_{3,WAG} \pm X\sigma_{3,WAG},\ x_{4,w} = \overline{x}_{4,WAG} \pm X\sigma_{4,WAG}, \\ x_{5,w} &= \overline{x}_{5,WAG} \pm X\sigma_{5,WAG},\ x_{8,w} = \overline{x}_{8,WAG} \pm X\sigma_{8,WAG}, \end{aligned} \tag{21}$$

The 96 single-phase simulations resulted in $16 \cdot 96 = 1536$ points to the ML model. In total, $1648 + 824 + 16 \cdot 96 = 4008$ data points were then applied in MOD2.

#### 2.3.3. Machine Learning Dataset Preparation

The 2472 WAG cases and 96 single-phase injection cases were both divided randomly between three sets: 70% in the training set, 15% in the validation set and the remaining 15% in the testing set [48]. Single-phase data points within each set were further split into 16 points, as described previously. See Table 1 for an overview of points in the models and datasets.

**Table 1.** Number and type of points in different datasets and models.

| MOD1 | Single Phase Cases | WAG Cases | Total |
|---|---|---|---|
| | 96 | 2472 | 2568 |
| MOD2 | Single phase cases | WAG cases | Total |
| Training (70%) | 68 × 16 = 1088 | 1730 | 2818 |
| Validation (15%) | 14 × 16 = 224 | 371 | 595 |
| Testing (15%) | 14 × 16 = 224 | 371 | 595 |
| Total | 96 × 16 = 1536 | 2472 | 4008 |

#### 2.3.4. Machine Learning Workflow

We apply LSSVM with radial basis kernel (RBK) function and either PSO, GWO, GA or GSA as optimizers. Each of the optimizers features its own strengths and disadvantages in finding global optima efficiently, as well as depending on its individual tuning parameters. They are swarm-based algorithms, making use of many potential solutions simultaneously and improving these solutions according to those performing best at a given iteration. The result is the existence of differently optimized LSSVM models, as indicated in Figure 2. Detailed explanations of the algorithms are provided in Appendices B and C, respectively. Each input parameter $x_i$ was normalized (denoted $x_N$) to a range between −1 and +1 based on the maximum and minimum values of the total dataset.

$$x_N = 2\frac{x - x_{min}}{x_{max} - x_{min}} - 1 \tag{22}$$

Assuming predefined values of the metaparameters $(\sigma, \gamma)$, LSSVM provides the function $y(x)$ and its coefficients $\alpha_k, b$ that minimize the error between the model predictions and observations of a given dataset (usually the training dataset) for those parameter choices [27]:

$$y(x) = \sum_{i=1}^{n} \alpha_i \exp\left(-\frac{||x_{Ni} - {x_N}^2||}{\sigma^2}\right) + b \tag{23}$$

For given metaparameters, the LSSVM algorithm is calibrated on the training set to provide choices of $\alpha_k, b$. The optimizer algorithm is used to search for the metaparameters $\sigma, \gamma$ that minimize the model prediction error on the validation dataset (i.e., many LSSVM models are calibrated on the training set and the one giving best prediction on the validation set is taken as the best). This systematically determines the best choice of metaparameters to avoid over- or under-fitting. The optimized LSSVM models were finally used to predict the data in the testing set. The model performing best overall was selected for further sensitivity analysis. For proper comparison, the optimizers were implemented with the same random initial solutions guesses (and velocities if applicable), search space and number of iterations. The optimizer parameters can be found in Table A4.

An advantage of LSSVM is its few (two) metaparameters and the automated optimization of its internal tuning parameters [27]. In comparison, artificial neural networks often need subjective selection of the number of nodes and layers and then comprehensive tuning of a vast number of weights and biases to train the network [49].

The correlation between input variables $x$ and output $y$ (RF) is quantified using Pearson correlation $r_{xy}^{P}$, Spearman rank $r_{xy}^{Sp}$ and distance correlation $r_{xy}^{D}$ coefficients. These indices,

respectively, indicate linear correlation, nonlinear monotonic correlation and nonlinear nonmonotonic correlation. The former two range between $-1$ and 1, while the latter ranges from 0 to 1. For all of them, 0 indicates no correlation. The goodness-of-fit between the calculated RF from MOD1 or MOD2 and the data values of RF were quantified using the coefficient of determination $R^2$ and the root mean square error RMSE. The definitions of the mentioned quantities are in Appendix D.

## 3. Results and Discussion

### *3.1. Preliminary Dataset Analysis*

The range and mean of the data points used in MOD1 and MOD2 (using $X = 0.5$) are listed in Table 2. The use of logarithms made the range of the different variables span a few units rather than orders of magnitude.

**Table 2.** Range of values for the total datasets used in MOD1 and MOD2 (using $X = 0.5$).

| **MOD1** | **Train** | | | **Val** | | | **Test** | | | **Tot** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Min | Mean | Max | Min | Mean | Max | Min | Mean | Max | Min | Mean | Max |
| $M^*$ | $-0.2$ | 1.4 | 3.4 | $-0.1$ | 1.5 | 3.2 | $-0.1$ | 1.4 | 3.2 | $-0.2$ | 1.4 | 3.4 |
| $y$ | 0.14 | 0.49 | 0.88 | 0.20 | 0.49 | 0.84 | 0.19 | 0.50 | 0.85 | 0.14 | 0.49 | 0.88 |
| **MOD2** | **Train** | | | **Val** | | | **Test** | | | **Tot** | | |
| | Min | Mean | Max | Min | Mean | Max | Min | Mean | Max | Min | Mean | Max |
| $x_1$ | 0 | 0.5 | 1 | 0 | 0.5 | 1 | 0 | 0.4 | 1 | 0 | 0.5 | 1 |
| $x_2$ | 0 | 0.5 | 1.1 | 0 | 0.5 | 1.1 | 0 | 0.4 | 1.1 | 0 | 0.5 | 1.1 |
| $x_3$ | 0 | 1.2 | 2.5 | 0 | 1.2 | 2.5 | 0 | 1.1 | 2.5 | 0 | 1.2 | 2.5 |
| $x_4$ | 0 | 1.3 | 3 | 0 | 1.3 | 3 | 0 | 1.4 | 3 | 0 | 1.3 | 3 |
| $x_5$ | 0.1 | 1.4 | 2.4 | 0.1 | 1.4 | 2.4 | 0.1 | 1.4 | 2.4 | 0.1 | 1.4 | 2.4 |
| $x_6$ | 0.0 | 1.4 | 2.3 | 0.0 | 1.4 | 2.3 | 0.0 | 1.3 | 2.3 | 0.0 | 1.4 | 2.3 |
| $x_7$ | $-4.6$ | $-2.6$ | $-0.9$ | $-4.6$ | $-2.7$ | $-0.9$ | $-4.6$ | $-2.7$ | $-0.9$ | $-4.6$ | $-2.6$ | $-0.9$ |
| $x_8$ | $-7.9$ | $-3.0$ | $-0.8$ | $-7.9$ | $-3.1$ | $-0.8$ | $-7.9$ | $-3.0$ | $-0.8$ | $-7.9$ | $-3.0$ | $-0.8$ |
| $y$ | 0.14 | 0.45 | 0.88 | 0.20 | 0.44 | 0.84 | 0.19 | 0.48 | 0.85 | 0.14 | 0.45 | 0.88 |

The Pearson, Spearman and distance correlation coefficients evaluated between RF and the input parameters were calculated for the two model datasets and are listed in Table 3. For MOD1, the two former coefficients were $\approx -0.94$ and the latter 0.93. Their magnitude being close to 1 indicates a strong linear correlation between RF and $x_0 = \log_{10} M^*$ and the negative sign indicates that a larger $M^*$ reduces RF.

For MOD2, Pearson correlation coefficients were reported for WAG cases only and for all cases when $X = 0.25$, 0.5 and 1 in Table 3. Spearman rank and distance correlation were calculated only for $X = 0.5$. Considering the WAG cases, several variables correlate with RF, especially $x_5$, $x_6$ which are the log of gas/oil and water/oil mobility ratios. They feature Pearson coefficients $r^P_{xy} \sim -0.5$ to $-0.6$ indicating that when they increase (less favorable mobility ratio towards the oil), RF is reduced. Heterogeneity, represented by $x_2$, also correlates with RF with a lower $r^P_{xy} \sim -0.27$, indicating that RF generally reduces when the heterogeneity factor increases. The hysteresis parameters $x_3, x_4$ correlate with RF in opposite ways to each other, with $r^P_{xy} \sim 0.15$ for $x_3$ and $-0.15$ for $x_4$. When $x_3$ (i.e., $\alpha$) increases, gas relative permeability is reduced and should improve RF. Higher $x_4$ (i.e., $\log C$) leads to less gas trapping and RF therefore decreases. The gravity numbers feature relatively poor Pearson correlation with RF, with $r^P_{xy} \sim 0.08$ for water-oil and 0.0045 for gas-oil. Similarly, the water volume fraction correlates little with RF, and $r^P_{xy} \sim -0.05$ is slightly negative. These results could be related to their coupled nature, as is discussed below.

**Table 3.** Pearson, Spearman and Distance correlation coefficients $r_{xy}$ evaluated for the total dataset between RF and the involved input parameters for MOD1 and MOD2.

| | **MOD1** | $r_{xy}^{P}$ | $r_{xy}^{Sp}$ | $r_{xy}^{D}$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $x_0$ | −0.94 | −0.95 | 0.93 | | | | | |
| | **MOD2** | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ |
| | WAG cases | −0.053 | −0.34 | 0.16 | −0.15 | −0.62 | −0.49 | 0.078 | 0.0045 |
| $r_{xy}^{P}$ | $X = 0.25$ | −0.055 | −0.27 | 0.11 | −0.10 | −0.58 | −0.49 | 0.059 | 0.0087 |
| | $X = 0.5$ | −0.055 | −0.27 | 0.10 | −0.099 | −0.57 | −0.48 | 0.057 | 0.0085 |
| | $X = 1$ | −0.055 | −0.27 | 0.087 | −0.083 | −0.53 | −0.45 | 0.053 | 0.0079 |
| $r_{xy}^{Sp}$ | $X = 0.5$ | −0.045 | −0.25 | 0.095 | −0.095 | −0.53 | −0.47 | 0.056 | 0.017 |
| $r_{xy}^{D}$ | $X = 0.5$ | 0.19 | 0.25 | 0.15 | 0.15 | 0.54 | 0.47 | 0.092 | 0.077 |

When considering the MOD2 datasets with single-phase data included for different $X$, we note that the magnitude and sign of the different Pearson coefficients are similar to when only the WAG cases were considered. The main difference is that the correlation is somewhat lower, especially the parameters with unspecified information during single-phase injection. This was expected, since we added points where RF does not vary with changes in these parameters.

When evaluating the dataset with Pearson rank correlation for $X = 0.5$, we observe similar, but slightly lower values as for the Pearson coefficient, except for $x_8$, where the correlation doubles, but remains very low. When considering the distance correlation coefficient, however, several input parameters correlate more strongly with recovery, indicating that their relation is nonmonotonic. In particular, the water fraction, $x_1$, features a higher distance correlation coefficient, of 0.19. WAG was expected to perform better than single-phase injection, with RF not changing linearly with $x_1$, but peaking. Gravity can be a cause of both low and improved sweep and the gravity numbers $x_7$, $x_8$ feature distance correlation coefficients around 0.08, where more impact is attributed the gas–oil gravity number in particular. Furthermore, the hysteresis parameters $x_3$, $x_4$ now seem to correlate more strongly. The three correlation coefficients are similar for $x_2$, indicating a relatively linear and monotic relation, so if all other parameters are constant, increased heterogeneity should reduce recovery.

Note that all the variables in MOD2 feature less correlation than the variable $x_0$ in MOD1, since they individually do not contain all the involved system parameters. The aim is for them to provide better predictions when combined.

### *3.2. Development of MOD1*

RF is plotted against $x_0$ in Figure 3 and demonstrates a clear correlation where higher $x_0$ gives lower RF. There is also significant scatter, meaning a given $x_0$ can be associated with a range of RF values. The data were fitted to a third-order polynomial function, given by the blue curve in Figure 3 and Equation (24). A higher order polynomial did not further reduce the RMSE, which means the remaining error was associated primarily with the scatter in the data.

$$\mathrm{RF} = \sum_{i=1:4} p_i (x_0)^{4-i}, \quad \left(R^2 = 0.889; \ \mathrm{RMSE} = 0.0498\right)$$
$$p_1 = 0.01645, \ p_2 = -0.06302, \ p_3 = -0.1393, \ p_4 = 0.7676 \tag{24}$$

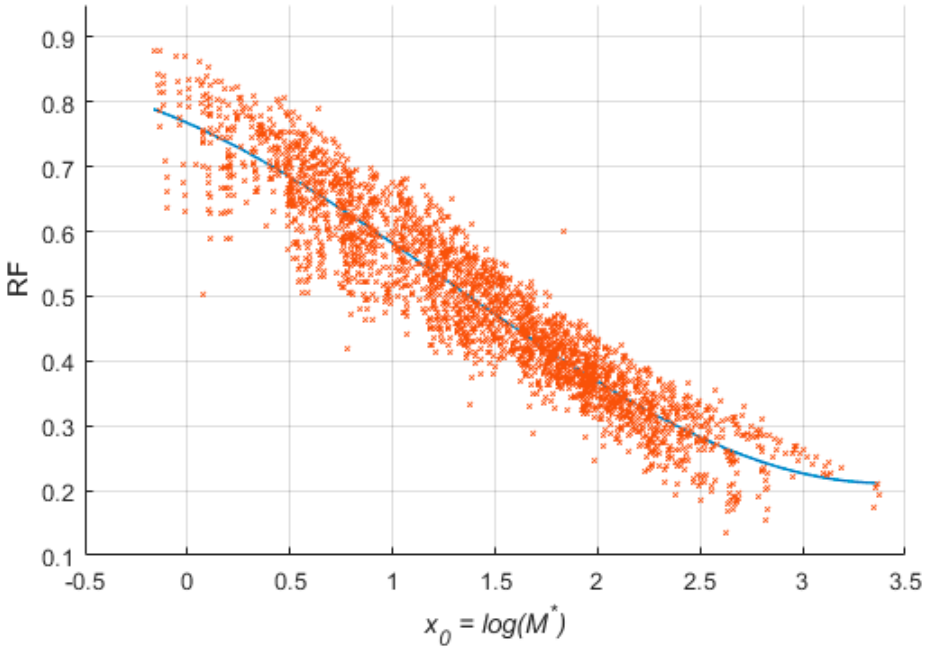


**Figure 3.** Datapoints plotted against corresponding values of $x_0$ for MOD1, defined using a third-order polynomial (blue line) of $x_0$.

The performance of the model is also shown by comparing the estimated RF and the data RF in Figure 4a together with a histogram, Figure 4b, of the residual errors (the difference between the estimated RF and the data point RF). The $R^2$ = 0.889 is relatively high. As seen in the histogram, the residuals are symmetrically distributed around zero and roughly 95% of the points estimate RF correctly within ±0.1. The RMSE, which can be considered a more typical error, is 0.050.

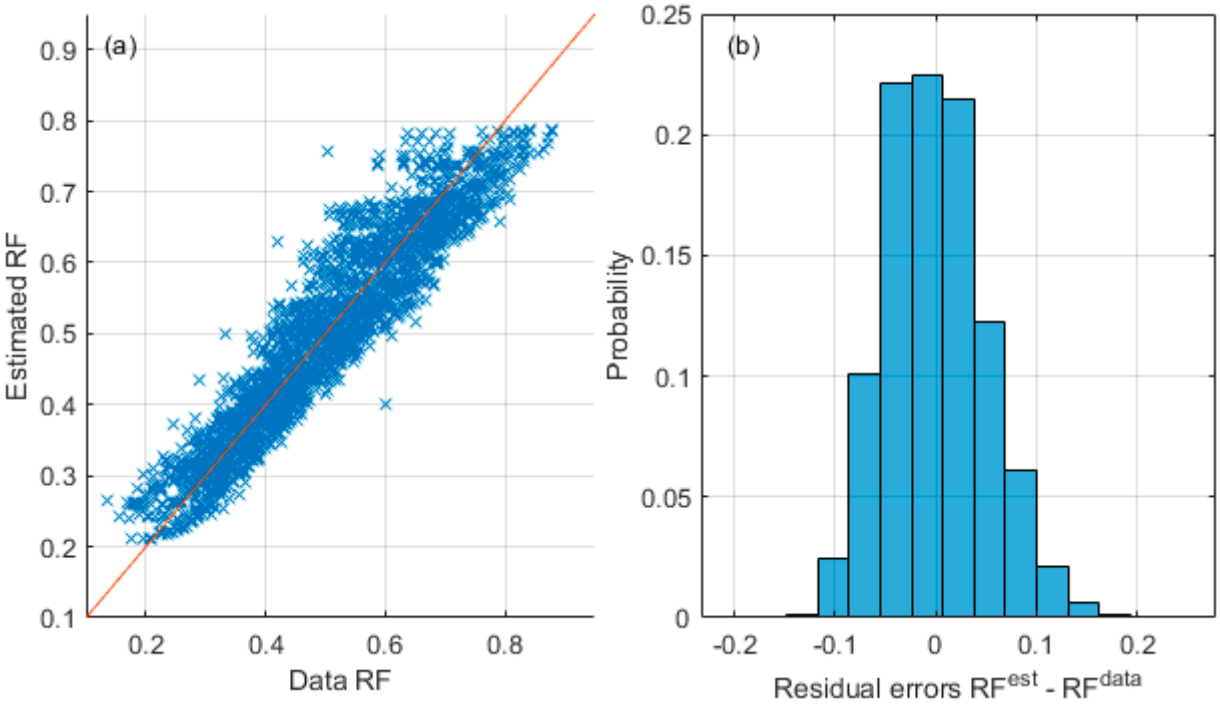


**Figure 4.** Comparison of estimated RF with MOD1 and actual datapoints (**a**) and a histogram of the residuals (**b**).

### 3.3. Development of LSSVM Model MOD2

The ML model MOD2 was developed using LSSVM and a dataset assuming $X = 0.5$. The best LSSVM model was determined using different approaches. First, a random choice of metaparameters $(\sigma, \gamma) = (1, 1)$ was used. Next, LSSVM was applied with the different optimization algorithms, PSO, GA, GWO and GSA, to systematically find the best metaparameters. As previously mentioned, for any combination of the metaparameters, LSSVM models were fitted to the training set and used to forecast the validation set. The metaparameters that resulted in the best performance in the validation set, after using a given optimizer algorithm, determined the best model. The test set was then forecasted.

In Figure 5, the performance of the different algorithms is illustrated as a function of the iterations performed. $R^2$ and RMSE for the validation set are plotted for the best solution at the given iteration, together with the corresponding values of $\log \gamma$ and $\log \sigma$ in plots a to d, respectively. The same initial solutions and number of iterations were applied in all the algorithms (different colors). Two different initializations were applied for robustness (dashed and full lines).

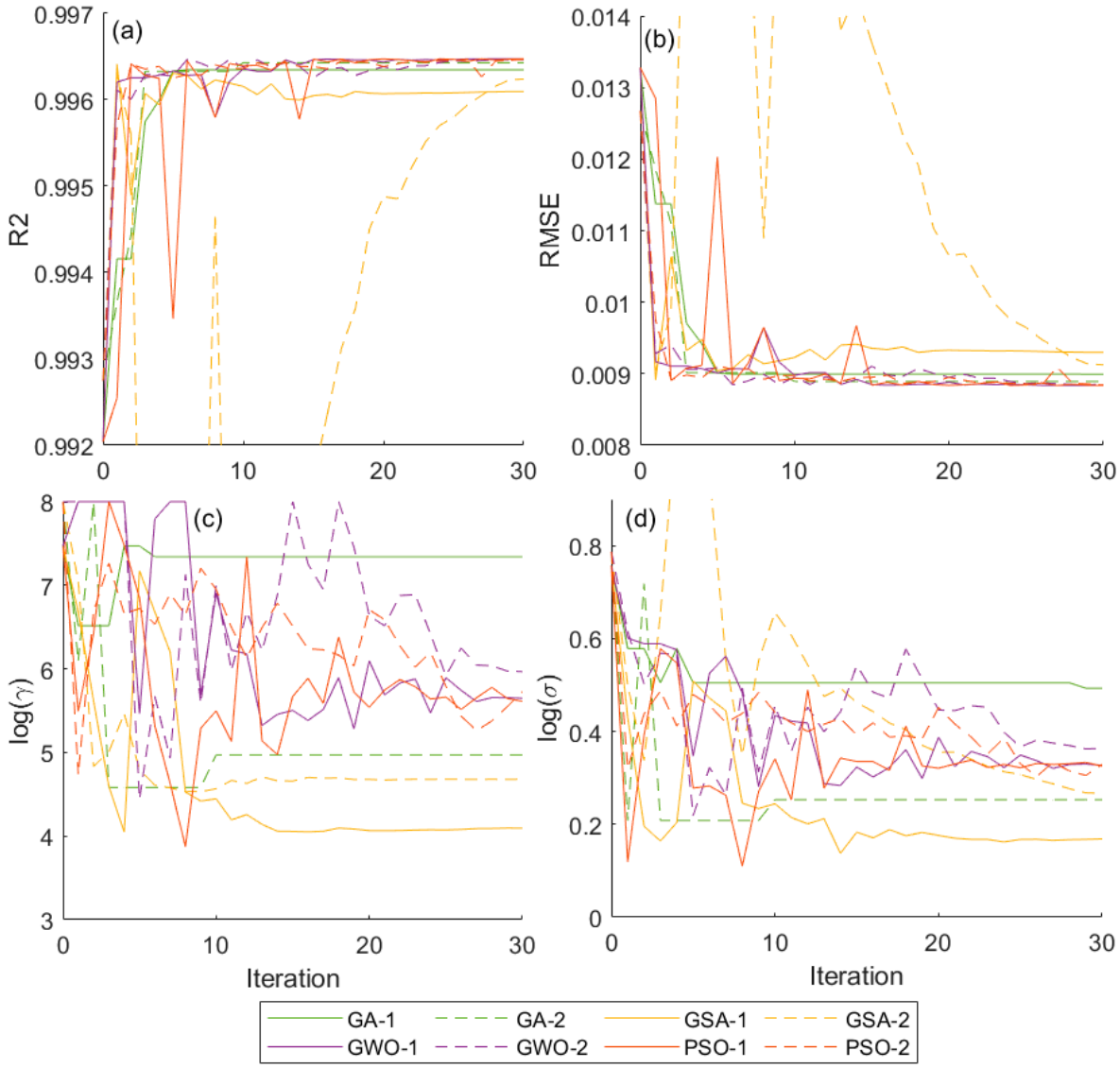


**Figure 5.** Illustration of optimizer performance in terms of the best solution's $R^2$ (**a**), RMSE (**b**) and search parameter values $log(\gamma)$ (**c**) and $log(\sigma)$ (**d**), at a given iteration. In total, 20 solutions were initiated and run for 30 iterations in each case. Two identical initializations (marked 1 and 2) were run for each algorithm.

In all cases, a high $R^2 \approx 0.996$ and low RMSE $\approx 0.009$ were obtained after 30 iterations for all the algorithms and both starting points, although GSA deviated from initially good solutions and converged slowly or to inferior solutions. Furthermore, GA seemed to not produce as good results as PSO and GWO. The two algorithms, PSO and GWO, exhibited very similar values of $\log \gamma \approx 5.5$ and $\log \sigma \approx 0.3$ and the lowest error indicating that they were better able to find the global optima. Notably, the values of $\log \gamma$ and $\log \sigma$ varied significantly during the iterations, but mostly exhibited very good performance. This may have been due to the ability of LSSVM to tune its internal parameters $\alpha_i$ and $b$ for any given $\gamma, \sigma$.

The best metaparameters obtained during the 30 iterations are listed in Table 4, considering all four algorithms and both initializations. The corresponding metrics ($R^2$ and RMSE) were calculated on the training, validation and testing sets. All the optimized models performed better than the algorithm with the arbitrarily preset metaparameters, although this choice also performed well, with $R^2$ greater than 0.969 on all three sets. The optimized models exhibited very consistent performance in the three datasets, with $R^2 \approx 0.999$ in the training set, $\approx$0.996 in the validation set and $\approx$0.992 on the testing set. The difference in $R^2$ was in the fourth digit for the first two sets and the third digit in the latter set. The RMSE was around 0.006 on the training set, 0.009 on the validation set and 0.015 on the test set, with GSA standing out with the highest RMSE. As final metaparameter values in the optimized model, we took an average of the four similar results from the PSO and GWO runs with two significant digits. Calculating the RMSE and $R^2$ metrics on the datasets confirmed that the performance with these values was still optimal (see Table 4). The LSSVM model with these parameters is referred to as MOD2 in what follows. Note especially that MOD2 is capable of predicting unseen single-phase data and thus accounts for the physics introduced during the modification of the dataset.

**Table 4.** Optimized LSSVM metaparameters using different optimizers and different initializations (marked 1 and 2) and corresponding performance metrics on the training, validation and testing datasets. The parameters used in the final model, MOD2, are indicated.

| | | | | **RMSE** | | | $R^2$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | Seed | $\log(\gamma)$ | $\log(\sigma)$ | Train | Val | Test | Train | Val | Test |
| LSSVM (preset) | | 0 | 0 | 0.0220 | 0.0202 | 0.0279 | 0.9821 | 0.9817 | 0.9691 |
| PSO-LSSVM | 1 | 5.6106 | 0.32535 | 0.0056 | 0.0088 | 0.0142 | 0.9988 | 0.9965 | 0.9920 |
| | 2 | 5.4335 | 0.30305 | 0.0055 | 0.0088 | 0.0142 | 0.9989 | 0.9965 | 0.9920 |
| GSA-LSSVM | 1 | 6.6812 | 0.42982 | 0.0058 | 0.0089 | 0.0150 | 0.9988 | 0.9964 | 0.9911 |
| | 2 | 7.0506 | 0.49871 | 0.0064 | 0.0090 | 0.0156 | 0.9985 | 0.9963 | 0.9904 |
| GWO-LSSVM | 1 | 5.6564 | 0.32883 | 0.0056 | 0.0088 | 0.0142 | 0.9988 | 0.9965 | 0.9919 |
| | 2 | 5.6698 | 0.32230 | 0.0055 | 0.0088 | 0.0144 | 0.9989 | 0.9965 | 0.9918 |
| GA-LSSVM | 1 | 7.3404 | 0.49280 | 0.0059 | 0.0090 | 0.0155 | 0.9987 | 0.9963 | 0.9904 |
| | 2 | 4.9708 | 0.25298 | 0.0054 | 0.0089 | 0.0140 | 0.9989 | 0.9964 | 0.9922 |
| Range (opt.) | | $\sim 2.4$ | $\sim 0.25$ | 0.0010 | 0.0002 | 0.0016 | 0.0004 | 0.0002 | 0.0018 |
| Final (MOD2) | | 5.6 | 0.32 | 0.0056 | 0.0088 | 0.0143 | 0.9988 | 0.9965 | 0.9919 |

The RMSE and $R^2$ were calculated with MOD2 for the total dataset as 0.0080 and 0.9976, respectively. These metrics are greatly improved compared to MOD1, which featured a corresponding RMSE of 0.0498 and an $R^2$ of 0.889. The calculated (with MOD2) and observed RF data are plotted against each other for the three datasets in Figure 6. For all three datasets, there is little scatter around the perfect match line. The residual errors were calculated for each datapoint in the full dataset and the results are plotted as a histogram

in Figure 7. Approximately 90% of the data feature errors in the estimated RF of less than 0.01, and 95% of the data feature errors less than 0.02.

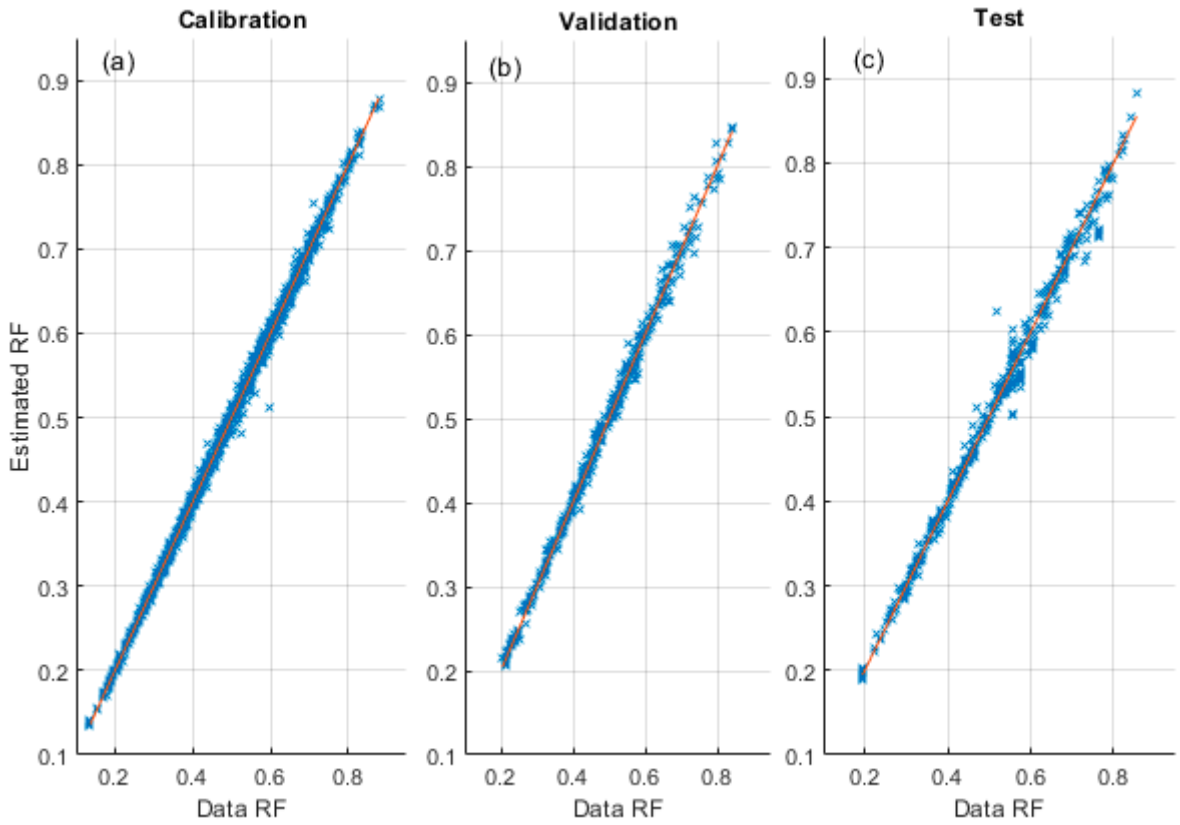


**Figure 6.** Comparison of estimated RF and actual datapoints on (**a**) the training set, (**b**) validation set and (**c**) test set. Estimated points are based on MOD2 (optimized LSSVM). The orange line represents perfect match.

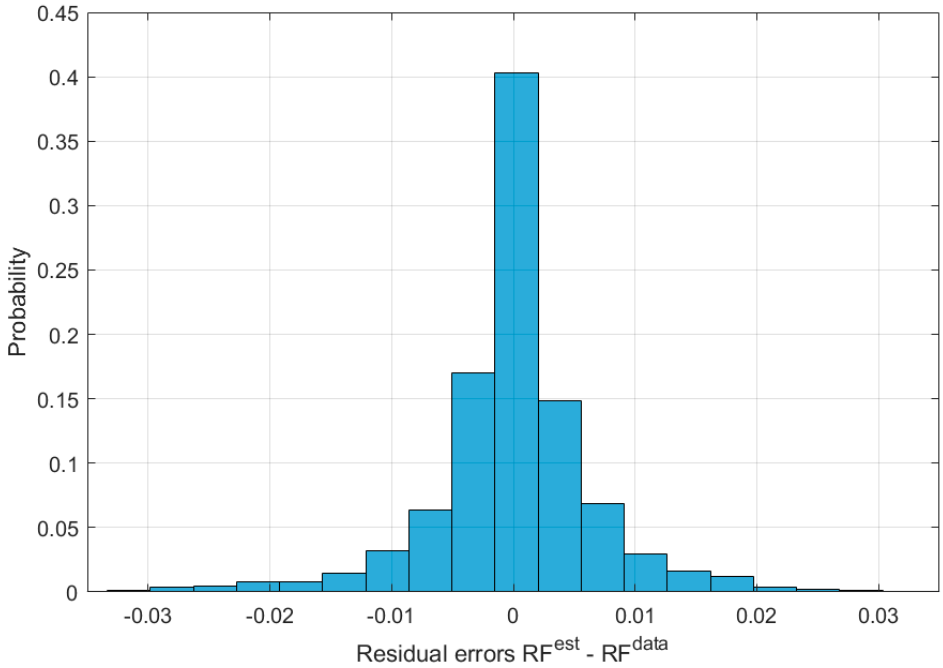


**Figure 7.** Histogram of residual errors (estimated RF minus actual RF ) for the total dataset based on MOD2 (optimized LSSVM model).

Partial derivatives with respect to each normalized variable, $\frac{\partial y}{\partial x_{Ni}}$, were calculated for each data point using MOD2 and histograms were created for each variable, as shown in Figure 8. The derivatives were calculated numerically and two choices of $\Delta x_N = 5 \cdot 10^{-2}$ and $10^{-3}$ were used. The two choices produced practically identical histograms, suggesting that the optimized LSSVM function did not suffer from oscillations (a sign of over-fitting). For each variable, a large fraction of the points featured positive and negative derivatives. Hence, changing the variable can affect RF positively or negatively, indicating coupling and room for finding optimal conditions.

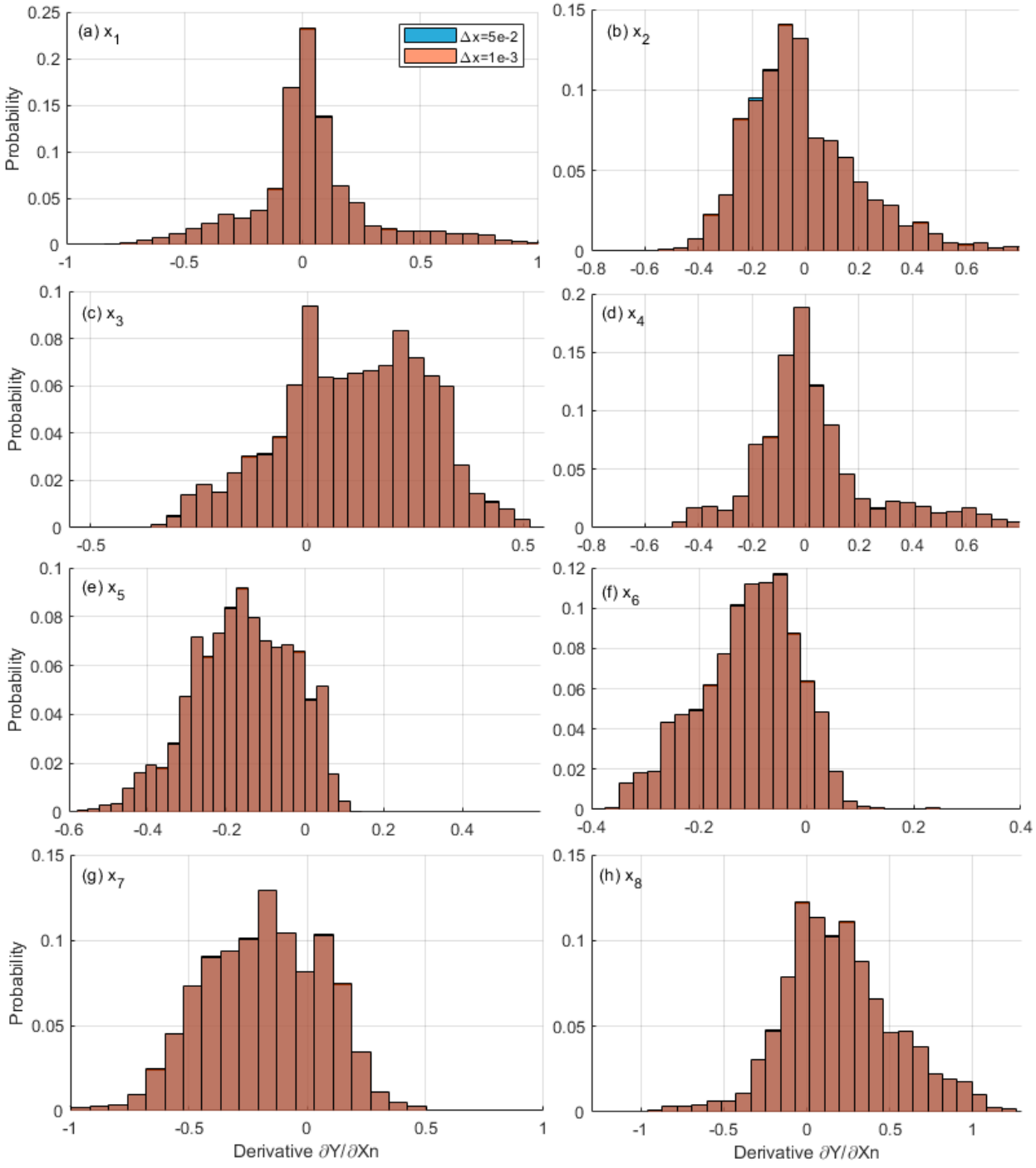


**Figure 8.** Histogram of partial derivatives for the total dataset based on MOD2 (the optimized LSSVM model). Each partial derivative is evaluated numerically with a small or large difference $\Delta x$.

For $\frac{\partial y}{\partial x_{N1}}$ we see positive and negative values, which is reasonable, since RF should be higher for WAG than single-phase injection. $\frac{\partial y}{\partial x_{N5}}$ and $\frac{\partial y}{\partial x_{N6}}$ are both dominated by negative values, since increasing the mobility ratio between gas and oil or between water and oil, respectively, should reduce RF. Higher water–oil gravity segregation is considered negative for RF, with $\frac{\partial y}{\partial x_{N7}}$ mainly negative. On the other hand, gas–oil gravity segregation is considered mainly positive for RF with a majority of points having $\frac{\partial y}{\partial x_{N8}}$ positive. This could be attributed to the better sweep of low-permeable layers in heterogeneous cases. Hysteresis appears to benefit recovery, as seen by a majority of positive $\frac{\partial y}{\partial x_{N3}}$, although the effect of $\frac{\partial y}{\partial x_{N4}}$ seems to be equally negative and positive.

### *3.4. Sensitivity Analyses with Optimized LSSVM Model MOD2*

The calibrated model, MOD2, was much better at predicting RF than MOD1 and was therefore pursued in the sensitivity analysis. Below, we present contour plots showing RF as a function of different input variables, while keeping the others constant. The parameters are kept within the total dataset range (see Table 2) in order to ensure model validity.

#### 3.4.1. Variation of Oil Viscosity

Oil viscosity can vary greatly from one reservoir to another. It proportionally impacts mobility ratios $M^*_{w/o}$ and $M^*_{g/o}$ in Equations (7) and (8), represented by $x_6$ and $x_5$. For low oil mobilities, the gravity numbers $N_G^{w/o}$ and $N_G^{g/o}$ (represented by $x_7$ and $x_8$) increase proportionally with oil viscosity but are less dependent if water or gas feature mobility that is similar to or lower than that of oil (see Equations (10) and (11)). For simplicity, we assume they are proportional. We vary the oil viscosity by 2.0 orders of magnitude, which is less than the smallest range of the four dimensionless numbers (2.3 for $x_6$), as seen in Table 2.

Four cases are defined in Table 5 with low or high heterogeneity (low or high $x_2$), and a low or high degree of hysteresis (low $x_3$ and high $x_4$ and opposite, respectively). For each of these cases, RF is plotted as a function of $x_1$ (the water fraction) and $x_5$ (representing the gas–oil mobility ratio) representing different viscosities (see Figure 9).From the figure, we observe that:

**Table 5.** Parameter selections for MOD2, where oil viscosity is varied and influences mobility ratios and gravity numbers. Four cases are considered according to heterogeneity and hysteresis.

| | **Low Het, Low Hyst** | **High Het, Low Hyst** | **Low Het, High Hyst** | **High Het, High Hyst** |
|---|---|---|---|---|
| $x_1$ | 0:1 | | | |
| $x_2$ | 0.25 | 1 | 0.25 | 1 |
| $x_3$ | 0 | 0 | 2.5 | 2.5 |
| $x_4$ | 3 | 3 | 0.5 | 0.5 |
| $x_5$ | 0.2:2.2 | | | |
| $x_6$ | 0.2:2.2 | | | |
| $x_7$ | −3:−1 | | | |
| $x_8$ | −3:−1 | | | |

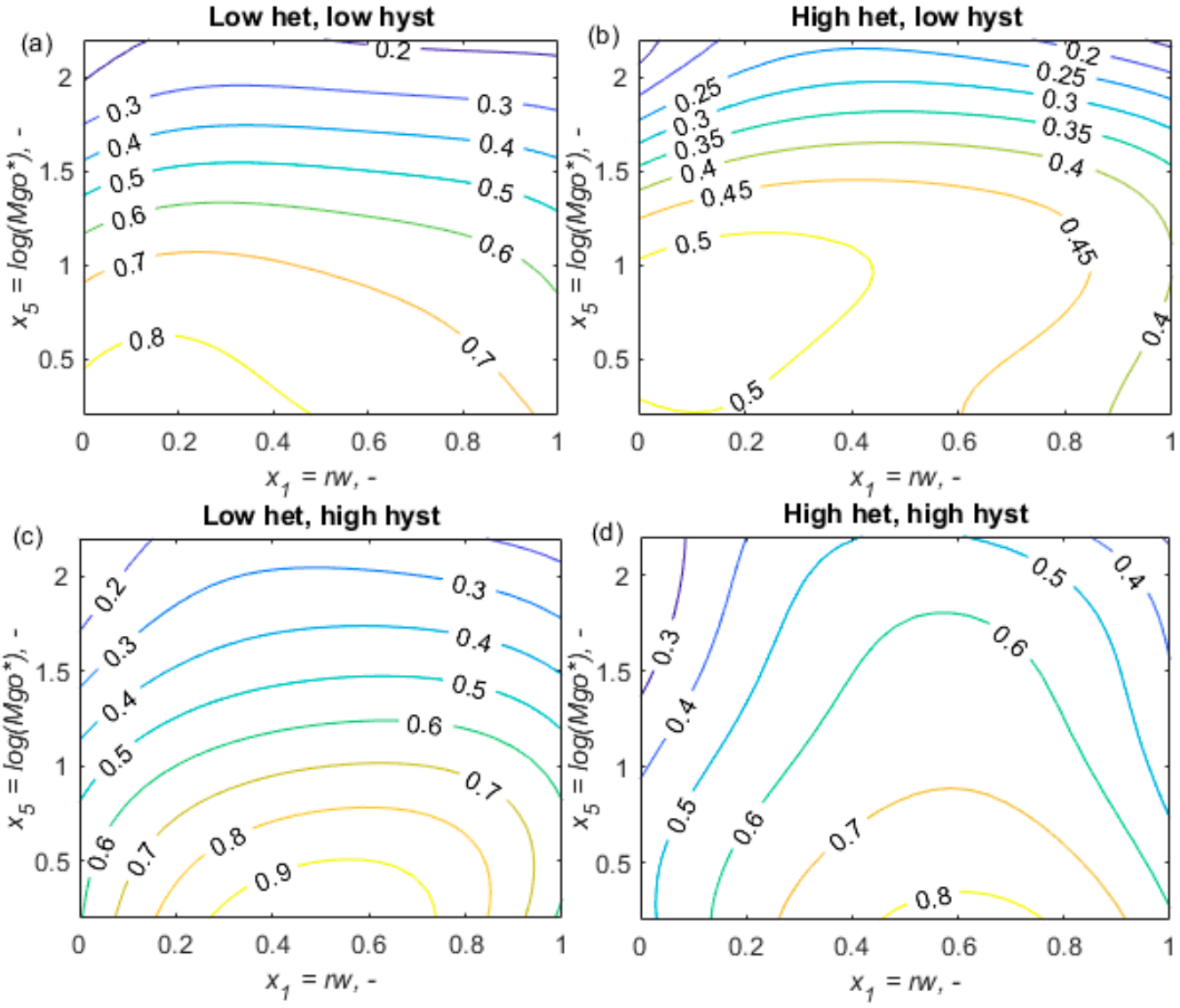


**Figure 9.** Contour plots of recovery factor RF plotted against $x_1 = r_w$ and $x_5 = log\left(M^*_{go}\right)$. The latter represents variation in oil viscosity, which affects all of $x_5, x_6, x_7, x_8$. The four cases are for low heterogeneity and hysteresis (**a**), high heterogeneity and low hysteresis (**b**), low heterogeneity and high hysteresis (**c**) and high heterogeneity and hysteresis (**d**). See all input values in Table 5.

- Optimal RF values were mainly obtained at an intermediate water fraction $0 < x_1$ (consider any line parallel with the x-axis), suggesting that WAG gives higher RF than single-phase injection. Cases with low hysteresis and favorable mobility ratios seem to give similar RF for water injection and WAG (although WAG with a low water fraction seems optimal) (see Figure 9a,b (low and high heterogeneity)).
- The advantage of WAG over single-phase injection was most clear when hysteresis was significant (see Figure 9c,d). The best water fraction produced RF up to 0.3 units higher than the worst fraction. This strong impact was mainly at low oil viscosity (low $x_5$) with optimal water fraction around 0.5–0.6. For higher oil viscosity or lower heterogeneity cases, WAG was in many cases only marginally better (~0.05 units) than the best single-phase injection.
- Increased oil viscosity reduced RF for a given water fraction (follow any line parallel with the y-axis). This was dominant over the WAG fraction at high viscosities, except for the highly heterogeneous cases with high hysteresis (Figure 9d). This demonstrates

the benefit of WAG in heterogeneous formations and that hysteresis is an important contributor.

- For a given heterogeneity (low or high), increased hysteresis improved RF (compare Figure 9c,d (high hyst) with Figure 9a,b (low hyst)). This was related to the improved gas–oil mobility ratio and reduced gravity segregation, which improves volumetric sweep. The optimal water fraction shifted to more central values, since both phases are needed for hysteresis.
- For a given hysteresis state, increased heterogeneity reduced RF, especially for cases with less viscous oil (compare Figure 9b,d (high het) with Figure 9a,c (low het)). For high hysteresis cases, increased heterogeneity increased RF in cases with more viscous oil.

To better understand the relation between viscosity, heterogeneity and hysteresis, we plotted RF as a function of $x_2 = logF_H$ and $x_5 = log\left(M^*_{go}\right)$ for $x_1 = 0.5$ (WAG injection with equal volume fractions of gas and water) for the two hysteresis cases in Figure 10. Each value of $x_5$ represents fixed oil viscosity and the curves cover the same viscosity range as before. We observed that that:

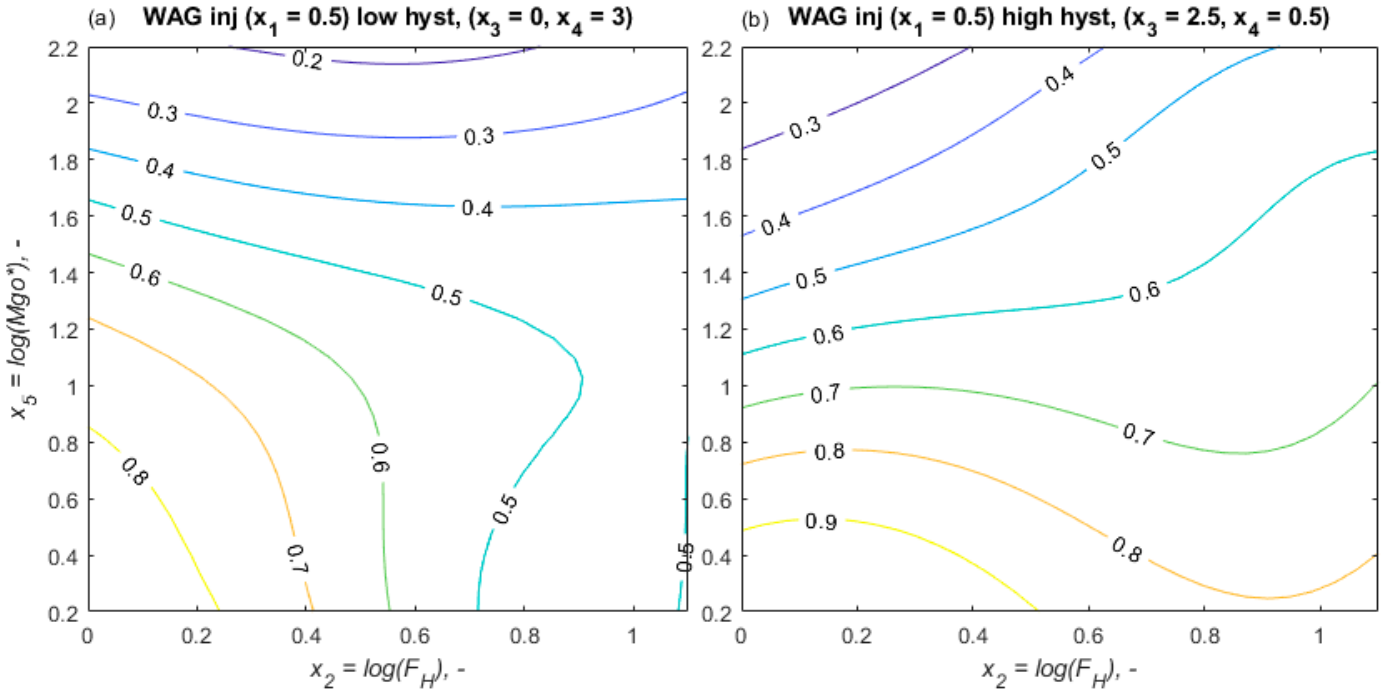


**Figure 10.** Contour plots of recovery factor RF plotted against $x_2 = log(F_H)$ and $x_5 = log\left(M^*_{go}\right)$. The latter represents variation in oil viscosity, which affects all of $x_5, x_6, x_7, x_8$. The cases are for WAG injection with $r_w = 0.5$ and either low (**a**) or high (**b**) hysteresis. See all input values in Table 5.

- For low hysteresis (Figure 10a), RF was very sensitive to heterogeneity for low oil viscosities and increased heterogeneity reduced RF. For high viscosity, RF changed little with heterogeneity.
- With significant hysteresis (Figure 10b), low-viscosity cases produced reduced RF at higher heterogeneity, while high-viscosity cases produced increased RF.

#### 3.4.2. Variation of Well Distance, Injection Rate or Density Difference

The distance between wells can vary from a dense pattern of a few hundred meters onshore to ~1000 m offshore. For fixed injection rates, a longer well distance $L_x$ proportionally increases the residence time and, hence, the gravity numbers (see (10)), represented by $x_7$ and $x_8$. Similarly, increasing the injection rates of water $Q_w$ and gas $Q_g$ equally reduces the residence times and the gravity numbers. Increased density differences reduce the

segregation time and increase the gravity numbers. If the height is varied but the injection rate is the same, we note that both segregation time and residence time change equally and there is no net change in the gravity numbers. Varying the aforementioned parameters does not affect the variables $x_1$ to $x_6$; we can thus investigate cases in which they are constant and only the gravity numbers change.

We plotted RF as a function of injected water fraction $x_1$ and log gravity number (equal values of $x_7$ and $x_8$). We investigated the role of mobility ratio, heterogeneity and hysteresis one by one. The different cases are listed in Table 6. The gravity numbers varied equally by 2.5 orders of magnitude.

**Table 6.** Parameter selections for MOD2 with cases demonstrating influence of gravity numbers according to heterogeneity, mobility ratio and hysteresis.

| | **Lo Het** | **Hi Het** | **Fav** | **Unfav** | **Lo Hyst** | **Hyst** |
|---|---|---|---|---|---|---|
| $x_1$ | 0:1 | | | | | |
| $x_2$ | 0 | 1 | 0.8 | | 0.3 | |
| $x_3$ | 1 | | 0 | | 0 | 2.5 |
| $x_4$ | 3 | | 3 | | 3 | 0 |
| $x_5$ | 2 | | 0.5 | 2 | 1.5 | |
| $x_6$ | 2 | | 0.5 | 2 | 1.5 | |
| $x_7$ | $-4$:$-1.5$ | | | | | |
| $x_8$ | $-4$:$-1.5$ | | | | | |

When low heterogeneity $x_2 = 0$ is considered (Figure 11a), RF stays fairly constant at low $N_g$ (when the impact from gravity is negligible) and decreases when $N_g$ is large due to gravity segregation and reduced vertical sweep. At high heterogeneity ($x_2 = 1$) in Figure 11b, RF is generally lower, but increases significantly with increases in the gravity number. Gravity therefore exerts a positive effect as more of the low-permeable layers are swept by gravity drainage into the highly permeable layers [47].

A relatively heterogeneous case ($x_2 = 0.8$ ) is considered where either mobility ratio is favorable, Figure 12a, or unfavorable, Figure 12b. In both cases, increased gravity number improves RF, but the effect is more pronounced in the favorable mobility ratio case. In the unfavorable mobility case, gravity exerts little impact until the gravity number exceeds $-3$. RF is generally higher in favorable mobility ratio cases compared to corresponding unfavorable mobility ratio cases.

In a relatively uniform case ($x_2 = 0.3$) with intermediate mobility ratios $x_5 = x_6 = 1.5$), hysteresis is varied. At low hysteresis, Figure 13a, increased gravity numbers increase RF moderately towards an optimal gravity number. At high hysteresis, Figure 13b, the optimal gravity number occurs at a lower value (for a given injected fraction). The peak can be related to improved sweep in low mobility layers, which becomes dominated by gravity segregation at the highest gravity numbers.

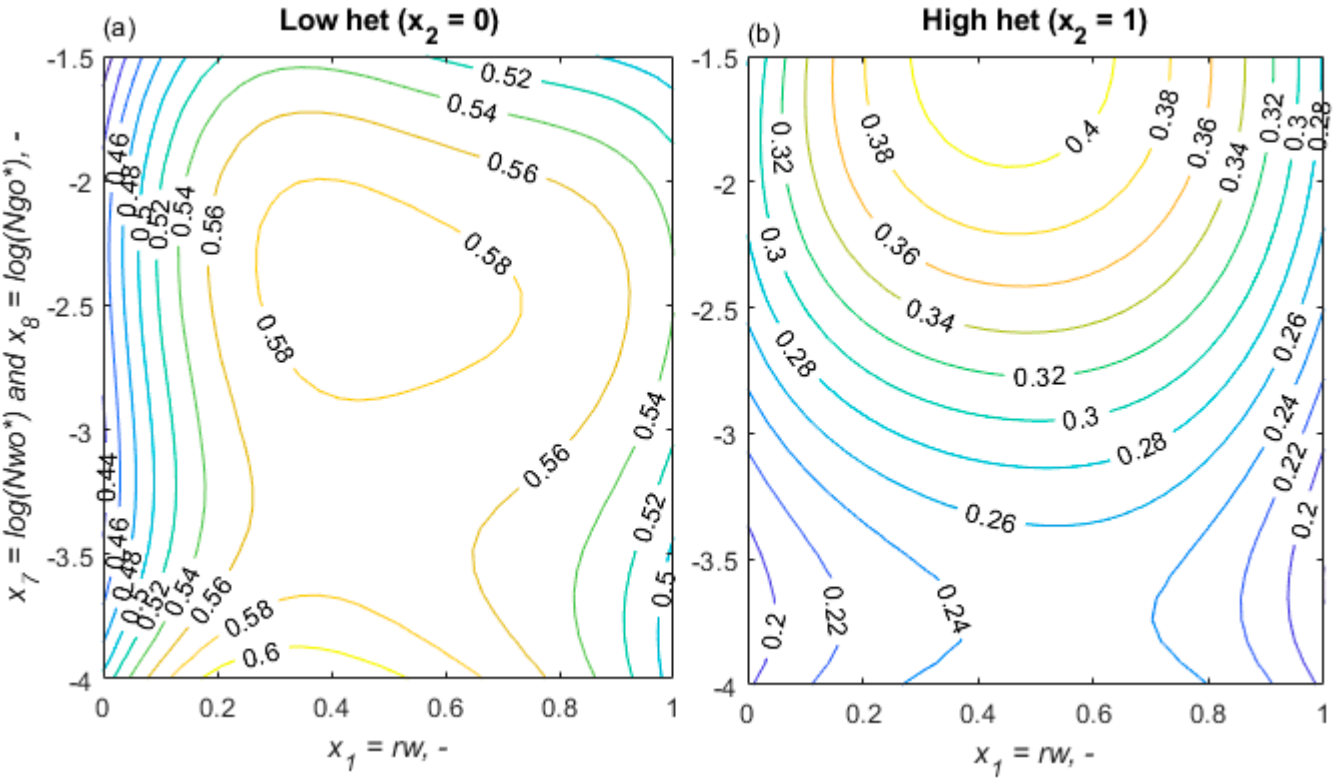


**Figure 11.** Contour plot of recovery factor RF plotted against $x_1 = r_w$ and log gravity number with equal values of $x_7$ and $x_8$. Low-heterogeneity (**a**) and high-heterogeneity (**b**) cases are shown (see all input values in Table 6).

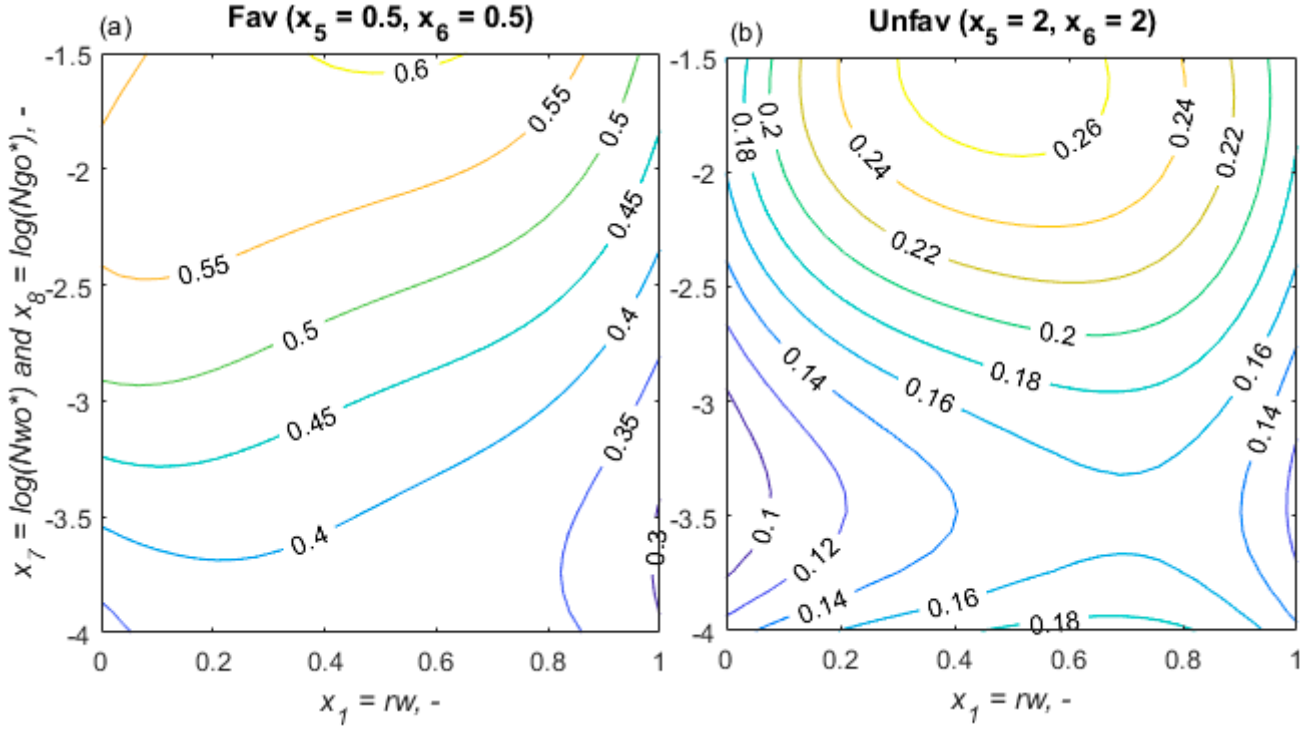


**Figure 12.** Contour plot of recovery factor RF plotted against $x_1 = r_w$ and log gravity number with equal values of $x_7$ and $x_8$. Favorable (**a**) and unfavorable (**b**) mobility ratio cases are shown (see all input values in Table 6).

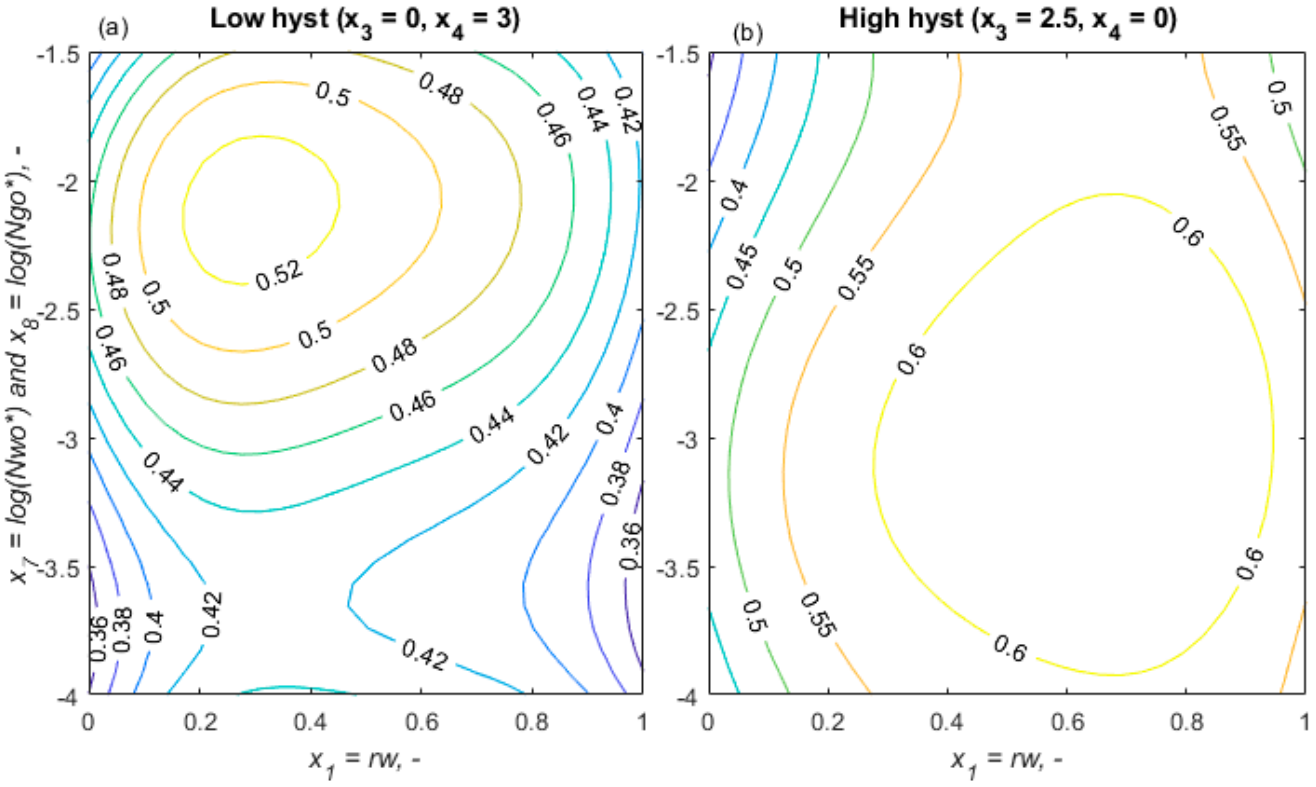


**Figure 13.** Contour plot of recovery factor RF plotted against $x_1 = r_w$ and log gravity number with equal values of $x_7$ and $x_8$. Low- (**a**) and high- (**b**) hysteresis cases are presented (see all input values in Table 6).

### 3.4.3. Handling Single Phase Data

The model MOD2 was trained to provide the same *RF* during single-phase injection when varying input variables related to hysteresis and the phase not injected (for example, gas during water injection). This was performed by generating points with different input values for parameters that should not exert an influence, but with the same output. To check how effectively this was captured by the calibrated model, we ran cases in which the hysteresis parameters $x_3, x_4$ and mobility ratio parameters $x_5, x_6$ were varied individually. RF was plotted against the relevant variable and $r_w$ ranging from gas to water injection. The input parameters are listed in Table 7 and the results are shown in Figure 14.

**Table 7.** Parameter selections for MOD2 cases to check response in going from multiphase to single-phase scenarios.

| | Vary $x_3$=$\alpha$ | Vary $x_4$=log$C$ | Vary $x_5$=log$M^*_{g/o}$ | Vary $x_6$=log$M^*_{w/o}$ |
|---|---|---|---|---|
| $x_1$ | | 0:1 | | |
| $x_2$ | | 0.8 | | |
| $x_3$ | 0:2.5 | 1 | 1 | 1 |
| $x_4$ | 1 | 0:3 | 1 | 1 |
| $x_5$ | 1.5 | 1.5 | 0.1:2.4 | 1.5 |
| $x_6$ | 1.5 | 1.5 | 1.5 | 0.0:2.3 |
| $x_7$ | | −2 | | |
| $x_8$ | | −2 | | |

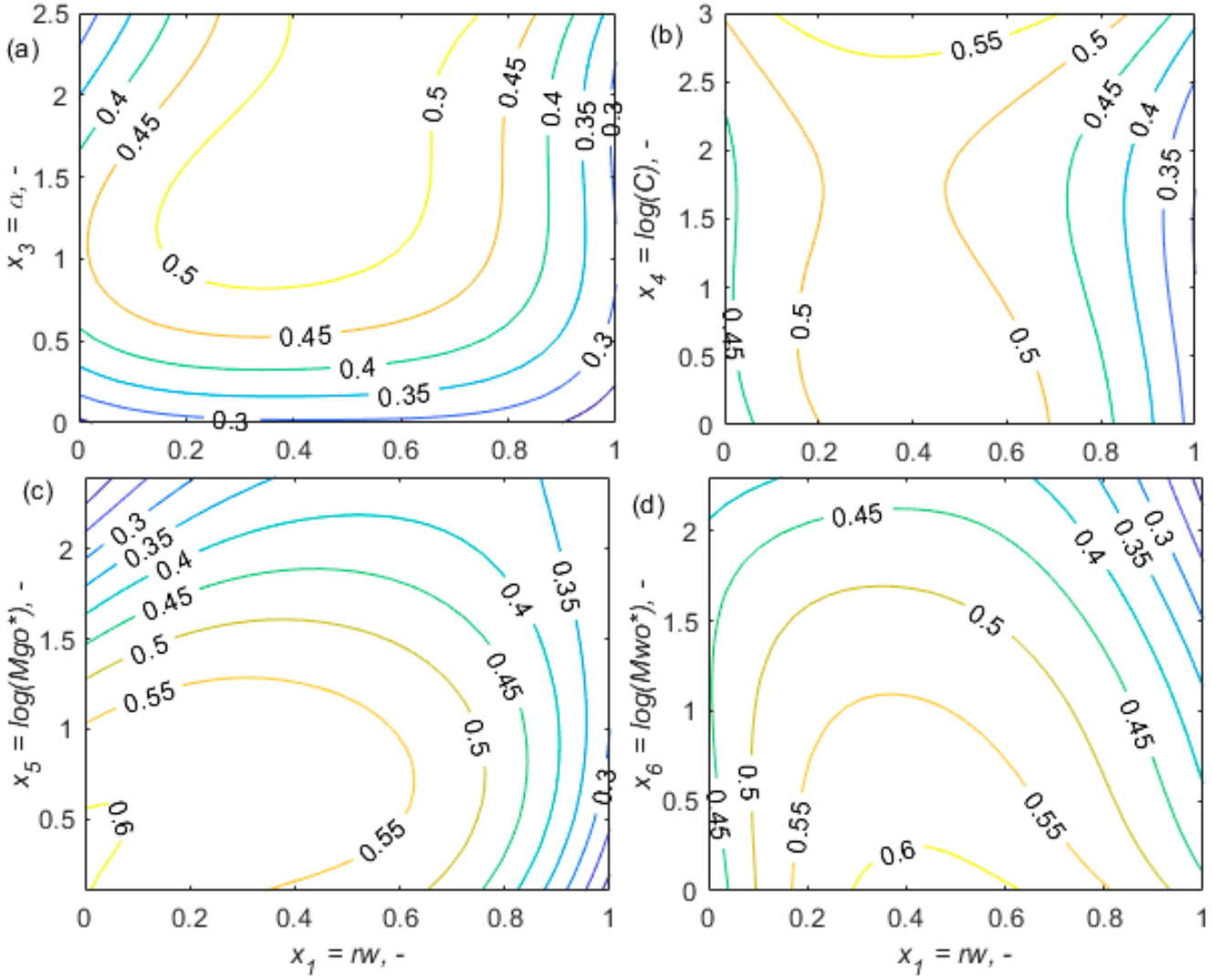


**Figure 14.** Contour plots of recovery factor RF as function of varying water fraction (horizontal axis) and the indicated parameter ($x_3$ in (**a**), $x_4$ in (**b**), $x_5$ in (**c**) and $x_6$ in (**d**)) on the vertical axis while holding other parameters fixed. $r_w = 0$ indicates gas injection and $r_w = 1$ water injection.

The variation of the hysteresis parameters $x_3$ and $x_4$, in Figure 14a,b, respectively, produces relatively constant RF with gas injection $r_w = 0$ (RF~0.45) and water injection $r_w = 1$ (RF~0.30), although the variation of $x_3$ during gas injection produces a wider range (RF = 0.30–0.45). The levels of RF differ, as is expected, since gas and water injection perform differently. Varying the gas-oil mobility ratio $M^*_{g/o}$ (via $x_5$ in Figure 14c) produces much less change in RF with water injection (RF~0.3) than gas or WAG injection. Similarly, varying the water–oil mobility ratio $M^*_{w/o}$ (via $x_6$ in Figure 14d) produces much less change in RF with gas injection (RF~0.45) than with water or WAG injection.

### *3.5. Application to a 3D Model*

The dataset used to train MOD1 and MOD2 was based on a 2D layered model. By obtaining effective parameters from 3D reservoir models, we can predict RF from MOD1 and MOD2. An artificial 3D heterogeneous model was considered with curvature, faults and three layers (see Figure 15). The average permeability, average porosity and layer thickness are listed in Table 8. The vertical permeability was half of the horizontal permeability. An injector and producer were placed at a distance of 1500 m and the pore volume was based on a width of 750 m. The RF was calculated after the injection of 1.5 PV, assuming five

injected fractions ($r_w$ equal 0, 0.33, 0.5, 0.67 and 1), two oil viscosities (30 and 110 cP) and low and high hysteresis (see values of $x_3$ and $x_4$).

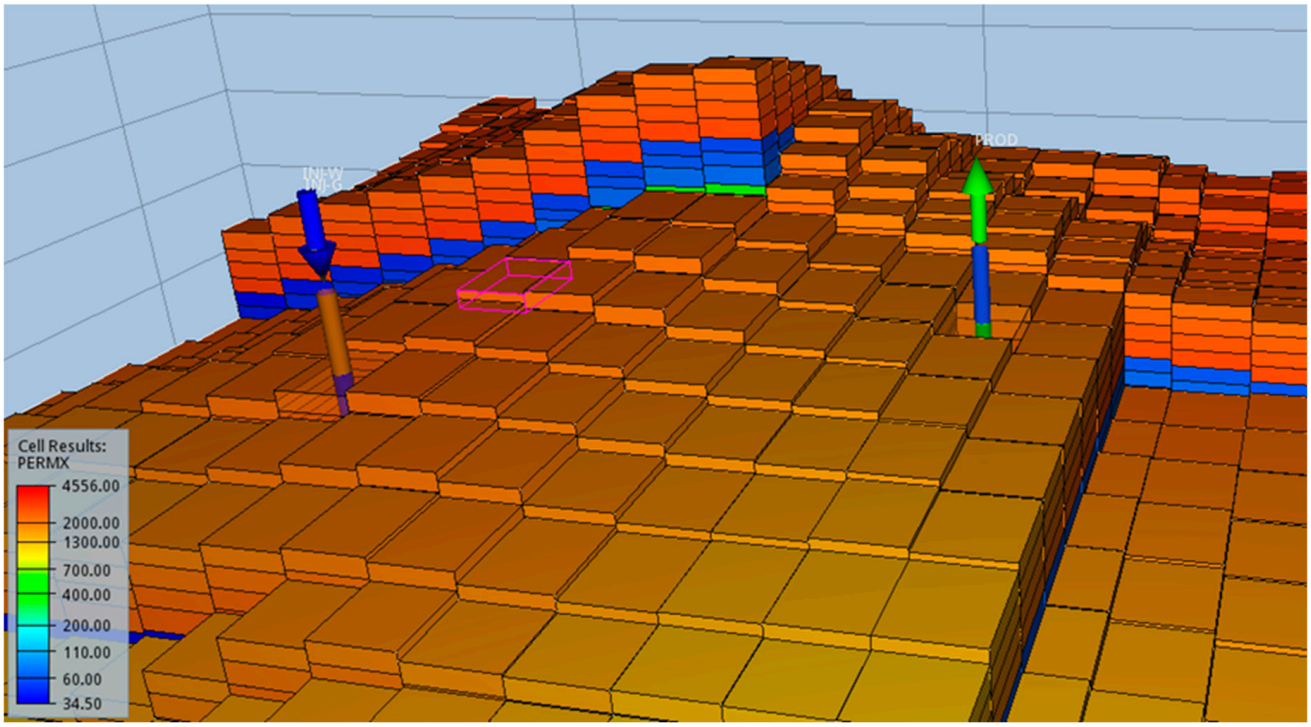


**Figure 15.** Illustration of the 3D model, where permeability and well placements are indicated.

**Table 8.** Input and calculated parameters for the 3D model. 'Low/hi' indicates degree of hysteresis in parameters $x_3$, $x_4$, while for parameters $x_5$ to $x_8$, values are calculated from two oil viscosities.

| **Layer** | $K_x$ **[mD]** | $\phi$ **[−]** | $h$ **[m]** | $x_1$ | **0, 0.33, 0.5, 0.67, 1** |
|---|---|---|---|---|---|
| 1 | 2170 | 0.324 | 42 | $x_2$ | 0.73 |
| 2 | 65.9 | 0.297 | 29 | $x_3(low, hi)$ | 0, 2.5 |
| 3 | 589 | 0.323 | 33 | $x_4(low, hi)$ | 3, 0 |
| | | | | $x_5(30, 110cP)$ | 1.83, 2.40 |
| $\Delta\rho_{wo}$ | 250 kg/m$^3$ | $V_p$ | $3.70 \times 10^7$ m$^3$ | $x_6(30, 110cP)$ | 1.50, 2.06 |
| $\Delta\rho_{go}$ | 450 kg/m$^3$ | $T$ | 20 years | $x_7(30, 110cP)$ | −1.28, −1.84 |
| $L$ | 1500 m | $s_{oi}$ | 0.84 | $x_8(30, 110cP)$ | −0.96, −1.52 |
| $W$ | 750 m | $Q$ | 7600 m$^3$/d | | |

In Figure 16, we plotted RF as a function of $x_1 = r_w$ for the four cases, calculated with the 3D model, MOD1 and MOD2. In these examples, RF with water injection (0.13 and 0.35 for high and low oil viscosity) is higher than with gas injection (0.09 and 0.22) and RF is lower with more viscous oil. When hysteresis is high, WAG exhibits the best performance out of all the models (the RF peaks at $0 < r_w < 1$). When hysteresis is low, the 3D model features similar RF for a high WAG fraction to water injection, while MOD1 indicates water injection as optimal and MOD2 still clearly supports WAG. MOD1 predicts a level of RF and change in RF that are more similar to those seen in the 3D model than MOD2. MOD2 predicts the level of RF in low hysteresis relatively well, but appears very sensitive to adding hysteresis. This is also seen through a difference in the single-phase injection RF for the same oil viscosity: about 0.2 units for water injection and for gas injection with low amounts of viscous oil, but, more reasonably, 0.03 units for gas injection with high amounts

of viscous oil. This indicates that this region of the model could be better calibrated. The water injection points in this case could be insufficiently near the single-phase points in the dataset. Furthermore, we do not expect MOD1 and MOD2 to predict the 3D model behavior identically as the geometries are not the same.

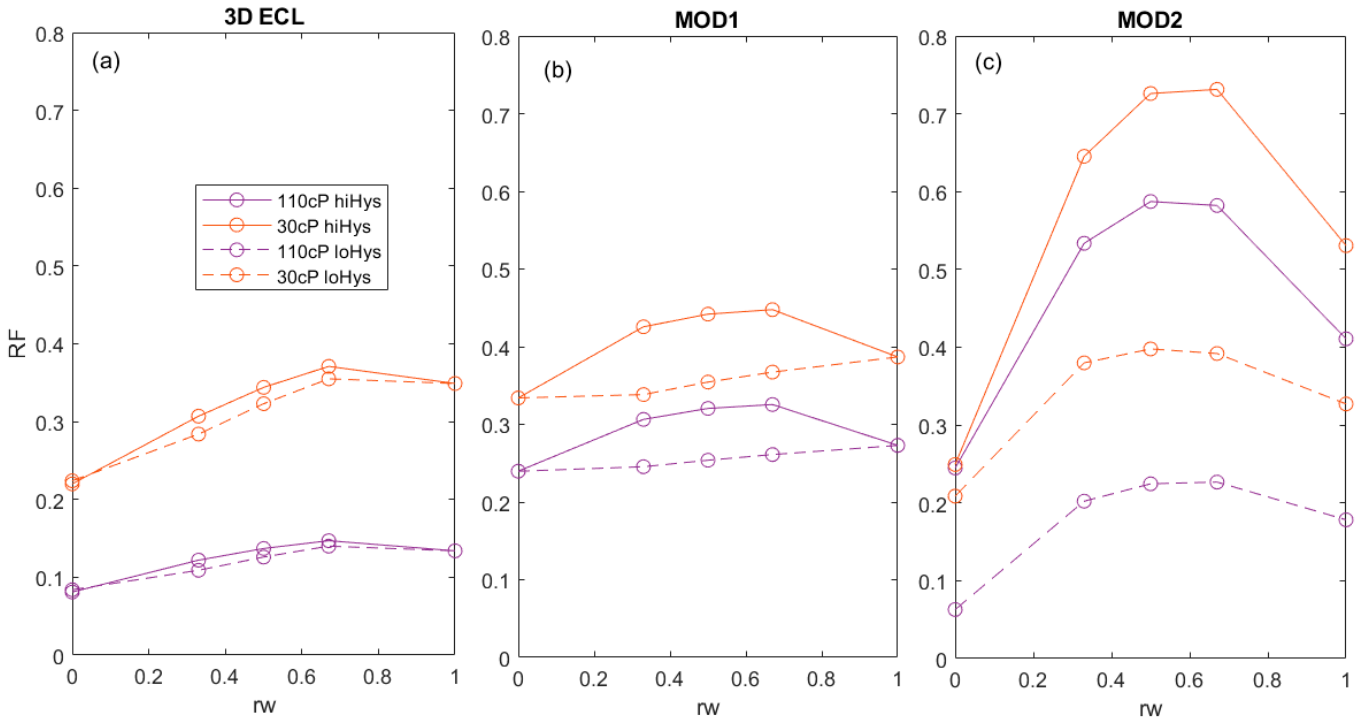


**Figure 16.** RF after 1.5 PV calculated for different injected WAG fractions ($r_w$), low or high oil viscosity and low or high degree of hysteresis, calculated based on a 3D Eclipse model (**a**), MOD1 (**b**) or MOD2 (**c**).

## 4. Conclusions

In this study, we interpreted a dataset of ~2500 points generated from single-phase and WAG injection reservoir model simulations to predict the recovery factor, RF. Two modeling approaches were selected. In MOD1, a universal dimensionless number $M^*$ derived from Nygård and Andersen [47] was selected as the single input variable and correlated by a polynomial expression. In MOD2, eight dimensionless numbers were used as input variables. Both choices included all the relevant input parameters to run the reservoir simulations. MOD2 was developed using LSSVM and optimized based on the best results from PSO, GA, GWO and GSA. The overall conclusions to this work can be summarized as follows:

- We demonstrated that it is possible to predict the recovery factor during single-phase and WAG injection.
- The LSSVM model optimized by GWO or PSO performed better than when optimized by GA or GSA.
- MOD2 with eight input variables clearly performed better than MOD1 with one input. Based on the total dataset, the RMSE and $R^2$ were 0.0080 and 0.998 for MOD2 and 0.050 and 0.889 for MOD1, respectively.
- The physics-based training of MOD2 was applied successfully. Single-phase injection data points were duplicated using different values in the input variables that should not affect RF, while keeping the same values for the relevant input variables and the output. The model correctly displayed little response to the irrelevant variables, but not for all conditions. Improvements could be made by adding more of these points or by training the model to include such constraints via an added penalty term in

the objective function. MOD1 was analytically independent of these variables during single-phase injection.

- Plotting histograms of partial derivatives of RF showed that for most input variables, increasing them would increase RF for some conditions, but reduce RF under others, demonstrating coupling in the data.
- The best model (MOD2) predicted that under identical conditions, an optimal injected WAG fraction existed that outperformed single-phase injection (water or gas). The benefit of WAG was much clearer when gas relative permeability hysteresis was significant.
- The mobility ratios were important input variables. Increased values tended to reduce RF.
- The roles of gravity numbers, heterogeneity and hysteresis were coupled. Strong gravity effects reduced RF in low-heterogeneity cases, but improved RF in heterogeneous cases.

Finally, some limitations of the study and recommendations should be mentioned. Some parameters were not varied in the dataset and their role can therefore not be predicted by the model. This includes the reservoir dip angle, capillary forces, starting WAG in tertiary mode (with some period of gas or water injection first), gas miscibility and tapering (changing WAG ratio with time). Furthermore, the heterogeneity of the model was mainly described by one parameter although porosity and vertical permeability appear in other dimensionless numbers. It could also matter how the heterogeneity appears, i.e., permeability increasing up or down. We note that several parameters that were not varied are included in a physically meaningful manner into dimensionless numbers that were varied. Thus, considering new values of Corey parameters, the vertical-to-horizontal permeability ratio and reservoir layer configurations are accounted for. The proposed methodology can be applied to predict the performance of other EOR techniques as well, but requires a similar development of representative dimensionless numbers and parameters capturing the EOR effect.

It is recommended to explore the potential of physics-based machine learning [50,51] in combination with dimensionless numbers describing complex systems, as was considered in this study. The methodology of modifying the dataset as described offers the advantage of applying ML algorithms in their standard form. On the downside, the dataset is enlarged and the physics are added around the specific datapoints, not as inherent part of the model.

**Author Contributions:** Conceptualization, P.Ø.A.; methodology, P.Ø.A.; software, P.Ø.A.; formal analysis, P.Ø.A.; investigation, P.Ø.A., J.I.N., A.K.; data curation, P.Ø.A., J.I.N., A.K.; writing—original draft preparation, P.Ø.A., A.K.; writing—review and editing, P.Ø.A.; visualization, P.Ø.A.; Supervision, P.Ø.A.; project administration, P.Ø.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The data can be provided upon request from the corresponding author.

**Acknowledgments:** Andersen acknowledges the Research Council of Norway and the industry partners, ConocoPhillips Skandinavia AS, Aker BP ASA, Vår Energi AS, Equinor ASA, Neptune Energy Norge AS, Lundin Norway AS, Halliburton AS, Schlumberger Norge AS and Wintershall DEA, of The National IOR Centre of Norway, for support.

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

## Nomenclature

*Roman*

| | |
|---|---|
| $b$ | LSSVM constant |
| $C$ | Land's trapping parameter |
| $F_H$ | Heterogeneity multiplier |
| $F_G$ | Gravity multiplier |

| | |
|---|---|
| $h_i$ | Layer height, m |
| $k_{ri}$ | Relative permeability |
| $k_{ri}^{max}$ | Relative permeability endpoints |
| $K_x$, $K_z$ | Horizontal and vertical absolute permeability, m |
| $L_x$ | Distance from injector to producer, m |
| $L_y$ | Width of reservoir, m |
| $L_z$ | Total height of reservoir, m |
| $M$ | Mobility ratio |
| $M^*_{g/o}, M^*_{\mathrm{w}/o}$ | Effective mobility ratios between gas-oil and between water-oil |
| $M^*$ | Effective three phase mobility ratio accounting for all mechanisms |
| $n_i$ | Corey exponents, |
| $N_p$ | Number of particles |
| $N_{it}$ | Number of PSO iterations |
| $N_G$ | Gravity number, - |
| $R^2$ | Coefficient of determination, - |
| $r_{xy}$ | Pearson correlation coefficient between vectors $x$ and $y$, - |
| $r_w$ | Water volume fraction in a WAG cycle, - |
| $RF$ | Recovery factor, - |
| $s_i$ | Phase saturation, - |
| $s_{ir}$ | Residual phase saturation, - |
| $t$ | Time, seconds |
| $x$ | Horizontal direction towards producer, m |
| $x_i$ | Input vector, - |
| $X$ | Standard deviance multiplier, - |
| $y$ | LSSVM output / RF, - |
| $z$ | Vertical direction downwards, m |
| *Greek* | |
| $\alpha$ | Carlson hysteresis parameter |
| $\alpha_i$ | LSSVM coefficients |
| $\gamma$ | Regularization coefficient |
| $\Delta\rho$ | Density difference, $\mathrm{kg/m^3}$ |
| $\lambda_i$ | Phase mobility $(\mathrm{Pa \cdot s})^{-1}$ |
| $\mu_i$ | Viscosity, Pa·s |
| $\rho_i$ | Phase density, $\mathrm{kg/m^3}$ |
| $\sigma$ | RBK width parameter |
| $\phi$ | Porosity |
| $\phi_1, \phi_2$ | Acceleration constants |
| $\omega$ | Damping factor |
| *Indices* | |
| * | characteristic value, |
| *arit* | arithmetic |
| *g* | gas |
| *G* | gravity |
| *harm* | harmonic |
| *i* | phase |
| *j* | layer |
| *o* | oil |
| *res* | residence |
| *init* | initial reservoir conditions |
| *seg* | segregation |
| *T* | total |
| *w* | water |
| *Abbreviations* | |
| EOR | Enhanced oil recovery |
| LSSVM | Least squares support vector machine |
| PSO | Particle swarm optimization |
| RMSE | Root mean square error |
| WAG | Water alternating gas |

## Appendix A. Reservoir Model Parameters

**Table A1.** Rock/grid properties and operational parameters. $N$ denotes number of cells in each direction, $L$ the respective lengths, $Q$ the volumetric rate.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $N_x$ | 100 | $L_x$ | 1000 m | $\Phi_j$ | 0.30 | $Q_w$ | 1014.6 $m^3/d$ | Half cycle duration | 45 d |
| $N_y$ | 1 | $L_y$ | 100 m | $h_j$ | 3 m | $Q_g$ | 1014.6 $m^3/d$ | Total injection volume, PVs | 1.5 PVs |
| $N_z$ | 81 | $L_z$ | 81 m | | | | | | |

**Table A2.** Reservoir flow properties in terms of relative permeability end points, Corey exponents, initial and residual saturations.

| | | | | | |
|---|---|---|---|---|---|
| $k_{row}^{max}$ | 0.25 | $n_{ow}$ | 2 | $S_{oi}$ | 0.842 |
| $k_{rog}^{max}$ | 0.25 | $n_{og}$ | 2 | $S_{wi} = S_{wr}$ | 0.158 |
| $k_{rw}^{max}$ | 0.05 | $n_w$ | 2 | $S_{gi} = S_{gr}$ | 0.00 |
| $k_{rg}^{max}$ | 0.005 | $n_g$ | 2 | $S_{orw}$ | 0.20 |
| | | | | $S_{org}$ | 0.10 |

**Table A3.** Specification of model heterogeneities. Each model had 9 layers with permeability distributed as specified. It was assumed that vertical and horizontal permeabilities were equal in each layer: $K_{z,j} = K_{x,j}$.

| | $K_x$ [mD] | | | |
|---|---|---|---|---|
| Layer 1 (top) | 300 | 300 | 500 | 1000 |
| 2 | 300 | 100 | 50 | 20 |
| 3 | 300 | 900 | 500 | 1000 |
| 4 | 300 | 300 | 50 | 20 |
| 5 | 300 | 100 | 500 | 1000 |
| 6 | 300 | 900 | 50 | 20 |
| 7 | 300 | 300 | 500 | 1000 |
| 8 | 300 | 100 | 50 | 20 |
| 9 (bottom) | 300 | 900 | 500 | 1000 |
| $F_H$ | 1.0 | 2.1 | 3.0 | 12.9 |

## Appendix B. Least Squares Support Vector Machines (LSSVM)

The support vector machine (SVM) algorithm was developed by Vapnik [25] and used to solve classification problems by building hyperplanes in multidimensional spaces that separated data into classes. Its application was extended to regression. The least squares support vector machine, or LSSVM, is a modification of SVM introduced by Suykens and Vandewalle [26]. The LSSVM regression algorithm is outlined below.

Consider a finite dataset with $n$ points $D = \{(x_1, y_1), \ldots\ldots, (x_n, y_n)\}$, where the input $x_i \in R^p$, $p$ being the number of input variables (in our case 1 or 8) and the output $y_i \in R$. The regression function is expressed as [27]:

$$f(x) = w^T \varphi(x) + b, \tag{A1}$$

where $\varphi$ is a higher dimensional function and $w$ is a weight vector that combines the contributions of each element of $\varphi$ to a scalar. Each output measurement $y_i$ is by definition equal to the regression plus the error $e_i$:

$$y_i = f(x_i) + e_i, \qquad (i = 1, \ldots, n) \tag{A2}$$

The LSSVM algorithm aims to minimize the objective function $J$ described as follows:

$$\min_{w,e} J(w,e) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^{n} e_i^2 \tag{A3}$$

$\gamma$ is called the regularization coefficient and its magnitude determines which of the two terms is minimized more. The error equations are treated as equality constraints:

$$y_i = w^T \varphi(x_i) + b + e_i, (i = 1, \ldots, n) \tag{A4}$$

Solving (A3) and (A4) simultaneously can be transformed to the problem of finding the saddle point of the Lagrange function $L$ which incorporates $J$ and the equality constraints:

$$L(w, b, e; \alpha) = J(w, e) - \sum_{k=1}^{n} \alpha_i \Big\{ w^T \varphi(x_i) + b + e_i - y_i \Big\} \tag{A5}$$

with Lagrange multipliers $\alpha_i$. The conditions for optimality are found by setting partial derivatives equal to zero:

$$\frac{dL}{dw} = 0 \rightarrow w = \sum_{i=1}^{n} \alpha_i \varphi(x_i) \tag{A6}$$

$$\frac{dL}{db} = 0 \rightarrow \sum_{i=1}^{n} \alpha_i = 0 \tag{A7}$$

$$\frac{dL}{de_i} = 0 \rightarrow \alpha_i = \gamma e_i, \ (i = 1, \ldots, n) \tag{A8}$$

$$\frac{dL}{d\alpha_i} = 0 \rightarrow w^T \varphi(x_i) + b + e_i - y_i = 0, \ (i = 1, \ldots, n) \tag{A9}$$

We can eliminate $e_i$ and $w$ from the above set of equations to obtain the remaining linear equations for $\alpha_i$ and $b$:

$$\sum_{i=1}^{n} \alpha_i = 0 \tag{A10}$$

$$b + \sum_{k=1}^{n} \alpha_k \varphi(x_k)^T \varphi(x_i) + \frac{1}{\gamma} \alpha_i = y_i \tag{A11}$$

By applying Mercer's condition, the product $\varphi(x_i)^T \varphi(x_j)$ is replaced by a kernel function $K(x_i, x_j)$:

$$\varphi(x_i)^T \varphi(x_j) = K(x_i, x_j), i, j = 1, \ldots, N (3.4) \tag{A12}$$

We can then solve for $\alpha_i$ and $b$ by solving the matrix form of (A10) and (A11):

$$\begin{bmatrix} 0 & 1 & \cdots & 1 \\ 1 & K(x_1, x_1) + \frac{1}{\gamma} & \ldots & K(x_1, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & K(x_n, x_1) & \cdots & K(x_n, x_n) + \frac{1}{\gamma} \end{bmatrix} \begin{bmatrix} b \\ \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} 0 \\ y_1 \\ \vdots \\ y_n \end{bmatrix} \tag{A13}$$

Combining (A4) with (A6) and (A12), the final form of the LSSVM regression function is given by:

$$y(x) = \sum_{i=1}^{N} \alpha_i K(x_i, x) + b \ (3.5) \tag{A14}$$

Different choices of kernel function can be made. The radial basis kernel (RBK) function was selected:

$$K(x_i, x) = \exp\left(-\frac{||x_i - x^2||}{\sigma^2}\right) \quad \text{(A15)}$$

$||x_i - x||$ denotes the Euclidian distance between vectors $x_i$ and $x$, while $\sigma$ is the width parameter.

The choice of the metaparameters $\sigma$ and $\gamma$ determines the LSSVM algorithm performance. $\sigma$ controls how rapidly the function can vary around the training data points $x_i$. For very small $\sigma$, the function equals the constant $b$ between the points $x_i$, while it matches $y_i$ at every $x_i$ of the training set. This results in the very poor prediction of new points. Large $\sigma$ linearizes the function (a straight line for a scalar input variable). Intermediate $\sigma$ are hence expected to capture non-linear trends. $\gamma$ controls how much weight is placed on minimizing the mismatch compared to minimizing the magnitude of the nonlinear terms. A very low $\gamma$ minimizes the coefficients of the nonlinear terms to zero and provides a constant function, equal to $b$. A very high $\gamma$ minimizes the mismatch between the function and the training set (between $y(x_i)$ and $y_i$) but allows it to be more nonlinear.

## Appendix C. Optimization Algorithms

Optimization algorithms are applied to find the optimal combination of LSSVM metaparameters, as represented by the vector $\beta = (\log_{10}\gamma, \log_{10}\sigma) \in R^2$. The applied parameters common and specific to the algorithms are listed in Table A4.

**Table A4.** Optimization algorithm parameters. No algorithm specific parameters were required for GWO.

| **Common** | | | **PSO** | | | **GA** | | |
|---|---|---|---|---|---|---|---|---|
| # particles /chromosomes/wolves | $N_p$ | 20 | Acceleration constants | $\phi_1$, $\phi_2$ | 1.5, 1.5 | Mutation rate | $\mu_r$ | 0.15 |
| # variables/genes | $N_{var}$ | 2 | Damping factor | $\omega$ | 0.8 | Mutation factor | $\mu_f$ | 0.1 |
| # iterations | $N_{it}$ | 30 | GSA | | | # elite chromosomes | $N_{elit}$ | 2 |
| Search range variable 1 | $\beta_1^{min}$, $\beta_1^{max}$ | −2, +8 | Initial gravity | $G_0$ | 2 | GWO | | |
| Search range variable 2 | $\beta_2^{min}$,$\beta_2^{max}$ | −3, +3 | Gravity reduction factor | $\alpha$ | 5 | - | | |
| Initial velocity range | | | Small constant | $\varepsilon$ | $10^{-4}$ | | | |

### *Appendix C.1. Particle Swarm Optimization (PSO)*

PSO was developed by Kennedy and Eberhart [52] and can be described as follows [53]:

a. Generate an initial set of $N_p$ 'particles', which are random solution vectors $\beta_n^0 (n = 1, \ldots, N_p)$, all in $R^2$. The entire set of particles is called the swarm.

$$\beta_{n,r}^0 = U_{n,r}\left(\beta_r^{min}, \beta_r^{max}\right), \quad (r = 1, 2). \quad \text{(A16)}$$

b. The indices $n$ and $r$ refer, respectively, to the particle and the parameter in the $n$ and $r$ vector while $U_{n,r}$ refers to the uniform probability distribution over the specified range.

The particles are assigned initial velocities $v_n^0 \in R^2$

$$v_{n,r}^0 = \frac{\beta_r^{max} - \beta_r^{min}}{\sqrt{N_p}} U_{n,r}(-1, 1), \quad (r = 1, 2). \quad \text{(A17)}$$

The initial velocity is set to be proportional to the search range and reduced by the number of particles, as they each can cover a shorter interval with more of them.

c. At a given iteration, the solution estimate of particle $n$ corresponds to its current 'position' in the search space, termed $\beta_n^{old}$. The quality of each of the $N_p$ solution estimates is evaluated by the coefficient of determination $R^2\left(\beta_n^{old}\right)$. The best solution position (with highest $R^2$) a particle obtains while it moves in the search space is saved and updated if it improves. These $N_p$ solution vectors are called $\beta_{n,opt}^p \ (n = 1, N_p)$. Similarly, the best solution of all the particles (the swarm) is termed $\beta_{opt}^s$. This position updates if the particles find a better solution.

d. New velocities $v_n^{new}$ are calculated for each particle $n$ based on the old velocity $v_n^{old}$ and how far the particle is from its historic best position $\beta_{n,opt}^p$ and from the swarm's historic best position $\beta_{opt}^s$:

$$v_n^{new} = \omega v_n^{old} + U(0,\phi_1)\left(\beta_{n,opt}^p - \beta_n^{old}\right) + U(0,\phi_2)\left(\beta_{opt}^s - \beta_n^{old}\right) \tag{A18}$$

$\phi_1$ and $\phi_2$ are acceleration constants, stating how quickly the particles steer towards the two currently best positions. A sum $\phi_1 + \phi_2 < 4$ avoids unbounded oscillation [53]. $\omega$ is a velocity damping factor. A value $\omega < 1$ refines searches at late iterations.

e. The position of each particle at the next iteration is updated by adding the velocity:

$$\beta_n^{new} = \beta_n^{old} + v_n^{new} \tag{A19}$$

Any particles exceeding the search space limits $\beta_r^{min}, \beta_r^{max}$ are adjusted to travel no farther than the limit.

f. Finally, the 'new' parameters are set as 'old' and a new iteration starts from point c. The procedure stops when a set number $N_{it}$ of iterations is completed.

*Appendix C.2. Gravitational Search Algorithm (GSA)*

GSA was developed by Rashedi et al. [54] and considers each solution as a particle.

a. Assign initial positions and velocities according to (A16) and (A17).

b. The gravitational constant is reduced from an initial value $G_0$ at iteration $t = 1$ according to a reduction factor $\alpha$ down to $G_0 \exp(-\alpha)$ at the last iteration:

$$G(t) = G_0 \exp\left(-\alpha \frac{t-1}{N_t - 1}\right). \tag{A20}$$

Calculate the relative fitness $m_n$ for each particle (here using $R^2$) at the current state.

$$m_n = \frac{\left(R^2\right)_n - \min\limits_n\left(R^2\right)}{\max\limits_n(R^2) - \min\limits_n(R^2)}. \tag{A21}$$

The 'mass' $M_n$ of each particle is then calculated as:

$$M_n = \frac{m_n}{\sum_{n=1:N_p} m_n}. \tag{A22}$$

c. For a given particle $n$, the force $F_{nj}$ working on it from another particle $j \neq n$ is given by:

$$F_{nj} = G \frac{M_n M_j}{|\beta_j - \beta_n| + \varepsilon}(\beta_j - \beta_n). \tag{A23}$$

where $|\beta_j - \beta_n|$ is the Euclidian distance between the particle positions and $\varepsilon$ is a small constant (to avoid division by zero).

d. The acceleration of particle $n$ is then its net force divided by the mass, where random weight components are introduced:

$$a_n^{old} = \frac{1}{M_n} \sum_{j=1:n} U_{n,j}(0,1) F_{nj}. \tag{A24}$$

e. The velocities and new positions are calculated as:

$$v_n^{new} = U(0,1) v_n^{old} + a_n^{old}. \tag{A25}$$

$$\beta_n^{new} = \beta_n^{old} + v_n^{new}. \tag{A26}$$

with coordinates limited by $\beta_r^{min}, \beta_r^{max}$. The procedure is repeated between steps b and e.

*Appendix C.3. Genetic Algorithm (GA)*

In GA each solution, $\beta_n \in R^2$ is called a chromosome and the individual elements $\beta_{n,r}$ $(r = 1, 2)$ are called the genes of the chromosome [55].

a. A first generation of chromosomes is initialized using (A16).
b. In 'Selection', pairs of two chromosomes from the previous generation, called parents, are combined to produce a new generation of chromosomes, 'children'. The selection of the parents is random with probability $P_n$ proportional to their relative fitness:

$$P_n = \frac{\mathrm{RMSE}_n^{-1}}{\sum_{n=1:N_p} \mathrm{RMSE}_n^{-1}}. \tag{A27}$$

c. 'Crossover' is then used to define the new generation chromosomes. In child 1 of a parent pair, the first gene is from parent 1 and the second gene from parent 2. For child 2 of that pair, the first gene is from parent 2 and the second from parent 1.

$$\beta_{c1,1}^{new} = \beta_{p1,1}^{old}, \ \beta_{c1,2}^{new} = \beta_{p2,2}^{old}, \ \beta_{c2,1}^{new} = \beta_{p2,1}^{old}, \ \beta_{c2,2}^{new} = \beta_{p1,2}^{old}. \tag{A28}$$

Generally, in problems with more than two genes, a crossover point must be defined to distinguish which genes are taken from which parent.

d. 'Mutation' is the operation of randomly modifying one or both genes in a child. The probability that a given gene is mutated is the mutation rate $0 \leq \mu_r \leq 1$. Thus, for the fraction $\mu_r$ of new genes we perform the following modification (while the rest $1 - \mu_r$ are not modified):

$$\beta_{n,r}^{new} \to \beta_{n,r}^{new} + \mu_f \left( \beta_r^{max} - \beta_r^{min} \right) U_{n,r}(-1,1) \tag{A29}$$

The factor $\mu_f$ is set to a low fraction so the mutation is low compared to the search range of the variables. The coordinates are limited by $\beta_r^{min}, \beta_r^{max}$.

e. 'Elitism' involves keeping some of the best chromosomes from the previous generation unmodified into the new generation.

*Appendix C.4. Grey Wolf Optimization (GWO)*

GWO was developed by Mirjalili et al. [56] and considers each solution $\beta_n \in R^2$ a 'wolf'.

a. Initialize the positions of the $N_p$ wolves according to (A16). In this algorithm, we call the positions $X$ instead of $\beta$.

b. At a given iteration the best, second-best and third-best solutions are called the alpha ($\alpha$), beta ($\beta$) and delta ($\delta$) wolves, respectively. The others are grouped as omega ($\omega$) wolves. The positions are denoted $X_\alpha$, $X_\beta$, $X_\delta$ and $X_\omega$, or $X$ for all the wolves.

c. Assume the 'prey' is located at a position $X_p$. A distance measure to the prey along coordinate $r$ is given by:

$$D_r = \left|C_r X_{p,r}(t) - X_r(t)\right|, \quad (r = 1:2) \tag{A30}$$

and the position at the next iteration is given as:

$$X_r(t+1) = X_{p,r}(t) - A_r D_r \tag{A31}$$

where the coefficients $C_r$ and $A_r$ are determined as follows:

$$A_r = (2U_r(0,1) - 1)a, \quad a = 2\left(1 - \frac{t-1}{N_{tot} - 1}\right), \quad C_r = 2U_r(0,1) \tag{A32}$$

The magnitude of $A_r$ makes it possible to move farther from the prey at early iterations (exploration) and closer at later iterations (exploitation). As the sign of $A_r$ can be positive or negative the new position can pass the prey on the given axis.

For each wolf, the position of the prey is estimated by the position of the three top wolves. The position at the next iteration is then the average of the three calculated new positions based on the top three wolves. Mathematically, this is expressed as:

$$D_{\alpha,r} = |C_{1,r} X_{\alpha,r} - X_r|, \quad D_{\beta,r} = \left|C_{2,r} X_{\beta,r} - X_r\right|, \quad D_{\delta,r} = |C_{3,r} X_{\delta,r} - X_r|, \tag{A33}$$

$$X_{1,r} = X_{\alpha,r} - A_{1,r} D_{\alpha,r}, \quad X_{2,r} = X_{\beta,r} - A_{2,r} D_{\beta,r}, \quad X_{3,r} = X_{\delta,r} - A_{3,r} D_{\delta,r}, \tag{A34}$$

$$X_r(t+1) = \frac{1}{3}(X_{1,r} + X_{2,r} + X_{3,r}), \quad (r = 1,2) \tag{A35}$$

The best position at a given iteration is described by the position of the alpha wolf. The coordinates are limited by $\beta_r^{min}$, $\beta_r^{max}$.

## Appendix D. Statistical Measures

Consider a dataset with $n$ points, in which we have a model trying to predict the observed output $y_i^{obs}$ but actually producing the modelled value $y_i^{mod}$ for point $i$. The goodness-of-fit the model provides for the dataset is quantified by the coefficient of determination $R^2$ between forecasted and true output values, also called the Nash–Sutcliffe efficiency [57,58]:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i^{obs} - y_i^{mod}\right)^2}{\sum_{i=1}^{n}\left(\overline{y}_i^{obs} - y_i^{mod}\right)^2}, \qquad \overline{y}_i^{obs} = \frac{1}{n}\sum_{i=1}^{n} y_i^{obs} \tag{A36}$$

where values from 0 to 1 correspond to no and perfect correlation, respectively. We also use the Root Mean Square Error (RMSE):

$$\mathrm{RMSE} = \left(\frac{1}{n}\sum_{i=1}^{n}\left(y_i^{obs} - y_i^{mod}\right)^2\right)^{0.5} \tag{A37}$$

Linear correlation between two variables, $x$ and $y$, is evaluated with the Pearson correlation coefficient $r_{xy}^P$:

$$r_{xy}^{\mathrm{P}} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}} \tag{A38}$$

A value close to $+1$ or $-1$ indicates strong positive or negative correlation, respectively. Nonlinear correlation is calculated using Spearman rank correlation $r_{xy}^{Sp}$. This is calculated by calculating the ranks of each value for the input variable $x$ and the output variable $y$, where the ranks $R(x_i), R(y_i)$ denote the position they would have sorted from least to largest. The rank correlation is then based on the covariance and standard deviations of these rank sets:

$$r_{xy}^{\mathrm{Sp}} = \frac{\mathrm{cov}(\mathrm{R}(x), R(y))}{std(R(x))std(R(y))} \tag{A39}$$

If none of the listed values are of equal rank, the above equation can be stated as

$$r_{xy}^{\mathrm{Sp}} = 1 - \frac{6\sum_{\mathrm{i}=1:n}(R(x_i) - R(y_i))^2}{n(n^2-1)} \tag{A40}$$

Nonmonotonic correlation can be detected using distance correlation [59]:

$$r_{xy}^{\mathrm{D}} = \frac{\mathrm{dCov}^2(x,y)}{\sqrt{dVar(x)dVar(y)}} \tag{A41}$$

defined by the fraction of squared distance covariance between $x$ and $y$ over the root mean of the distance variances of $x$ and $y$, respectively. We refer to the original work for more details.

## References


1. Stenmark, H.; Andfossen, P.O. Snorre WAG Pilot—A Case Study. In Proceedings of the IOR 1995—8th European Symposium on Improved Oil Recovery, Vienna, Austria, 15–17 May 1995.
2. Christensen, J.R.; Stenby, E.H.; Skauge, A. Review of WAG Field Experience. *SPE Reserv. Eval. Eng.* **2001**, *4*, 97–106. [CrossRef]
3. Sadik-Zada, E.R.; Loewenstein, W. A Note on Revenue Distribution Patterns and Rent-Seeking Incentive. *Int. J. Energy Econ. Policy* **2018**, *8*, 196–204.
4. Afzali, S.; Rezaei, N.; Zendehboudi, S. A comprehensive review on Enhanced Oil Recovery by Water Alternating Gas (WAG) injection. *Fuel* **2018**, *227*, 218–246. [CrossRef]
5. Sanchez, N.L. Management of water alternating gas (WAG) injection projects. In Proceedings of the Latin American and Caribbean Petroleum Engineering Conference, Caracas, Venezuela, 21–23 April 1999.
6. Sohrabi, M.T.D.H.; Tehrani, D.H.; Danesh, A.; Henderson, G.D. Visualization of oil recovery by water-alternating-gas injection using high pressure micromodels. *SPE J.* **2004**, *9*, 290–301. [CrossRef]
7. Chen, B.; Reynolds, A.C. Ensemble-based optimization of the water-alternating-gas-injection process. *SPE J.* **2016**, *21*, 786–798. [CrossRef]
8. Green, D.W.; Willhite, G.P. *Enhanced Oil Recovery*, 2nd ed.; Henry, L., Ed.; Society of Petroleum Engineers: Richardson, TX, USA, 2018.
9. Kulkarni, M.M.; Rao, D.N. Experimental investigation of miscible and immiscible Water Alternating Gas (WAG) process performance. *J. Pet. Sci. Eng.* **2005**, *48*, 1–20. [CrossRef]
10. Andersen, P.Ø. A simplified modelling approach for petroleum recovery by spontaneous imbibition in naturally fractured reservoirs. *J. Nat. Gas Sci. Eng.* **2019**, *63*, 95–114. [CrossRef]
11. Andersen, P.Ø. Early- and Late-Time Analytical Solutions for Cocurrent Spontaneous Imbibition and Generalized Scaling. *SPE J.* **2021**, *26*, 220–240. [CrossRef]
12. Land, C.S. Calculation of Imbibition Relative Permeability for Two- and Three-Phase Flow From Rock Properties. *Soc. Pet. Eng. J.* **1968**, *8*, 149–156. [CrossRef]
13. Stone, H. Estimation of Three-Phase Relative Permeability And Residual Oil Data. *J. Can. Pet. Technol.* **1973**, 12. [CrossRef]
14. Baker, L.E. Three-phase relative permeability correlations. In Proceedings of the SPE Enhanced Oil Recovery Symposium, Tulsa, OK, USA, 16–21 April 1988.
15. Carlson, F.M. Simulation of relative permeability hysteresis to the nonwetting phase. In Proceedings of the SPE Annual Technical Conference and Exhibition, San Antonio, TX, USA, 4–7 October 1981.
16. Larsen, J.A.; Skauge, A. Methodology for numerical simulation with cycle dependent relative permeabilities. *SPE J.* **1998**, *3*, 163–173. [CrossRef]
17. Spiteri, E.J.; Juanes, R. Impact of relative permeability hysteresis on the numerical simulation of WAG injection. *J. Pet. Sci. Eng.* **2006**, *50*, 115–139. [CrossRef]
18. Mahzari, P.; Sohrabi, M. An improved approach for estimation of flow and hysteresis parameters applicable to WAG experiments. *Fuel* **2017**, *197*, 359–372. [CrossRef]

19. Bourgeois, M.; Joubert, T.; Dominguez, V. Analysis of 3-phase Behavior in WAG Injections for Various Wettabilities. In Proceedings of the IOR 2019—20th European Symposium on Improved Oil Recovery, Pau, France, 8–11 April 2019; pp. 1–16. [CrossRef]
20. Cheng, G.; Guo, R.; Wu, W. Petroleum Lithology Discrimination Based on PSO-LSSVM Classification Model. In Proceedings of the 2010 Second International Conference on Computer Modeling and Simulation, Sanya, China, 22–24 January 2010; Volume 4, pp. 365–368.
21. Alvarado, V.; Ranson, A.; Hernandez, K.; Manrique, E.; Matheus, J.; Liscano, T.; Prosperi, N. Selection of EOR/IOR opportunities based on machine learning. In Proceedings of the European Petroleum Conference, Aberdeen, UK, 29 October 2002.
22. Tahmasebi, P.; Javadpour, F.; Sahimi, M. Data mining and machine learning for identifying sweet spots in shale reservoirs. *Expert Syst. Appl.* **2017**, *88*, 435–447. [CrossRef]
23. Chamkalani, A.; Zendehboudi, S.; Bahadori, A.; Kharrat, R.; Chamkalani, R.; James, L.; Chatzis, I. Integration of LSSVM technique with PSO to determine asphaltene deposition. *J. Pet. Sci. Eng.* **2014**, *124*, 243–253. [CrossRef]
24. Amar, M.N.; Ghriga, M.A.; Ouaer, H.; Seghier, M.E.A.B.; Pham, B.T.; Andersen, P. Østebø Modeling viscosity of $CO_2$ at high temperature and pressure conditions. *J. Nat. Gas Sci. Eng.* **2020**, *77*, 103271. [CrossRef]
25. Vapnik, V. The nature of statistical learning theory. Springer science & business media, New York, 1999. Springer Science & Business Media: New York, NY, USA, 1999.
26. Suykens, J.A.K.; Vandewalle, J. Least Squares Support Vector Machine Classifiers. *Neural Process. Lett.* **1999**, *9*, 293–300. [CrossRef]
27. Suykens, J.A.; Van Gestel, T.; De Brabanter, J. *Least Squares Support Vector Machines*; World Scientific: Singapore, 2002.
28. Alizadeh, S.M.; Alruyemi, I.; Daneshfar, R.; Mohammadi-Khanaposhtani, M.; Naseri, M. An insight into the estimation of drilling fluid density at HPHT condition using PSO-, ICA-, and GA-LSSVM strategies. *Sci. Rep.* **2021**, *11*, 1–14. [CrossRef] [PubMed]
29. Bian, X.-Q.; Song, Y.-L.; Mwamukonda, M.K.; Fu, Y. Prediction of the sulfur solubility in pure H2S and sour gas by intelligent models. *J. Mol. Liq.* **2020**, *299*, 112242. [CrossRef]
30. Mokarizadeh, H.; Atashrouz, S.; Mirshekar, H.; Hemmati-Sarapardeh, A.; Pour, A.M. Comparison of LSSVM model results with artificial neural network model for determination of the solubility of $SO_2$ in ionic liquids. *J. Mol. Liq.* **2020**, *304*, 112771. [CrossRef]
31. Ouaer, H.; Hosseini, A.H.; Amar, M.N.; Seghier, M.E.A.B.; Ghriga, M.A.; Nabipour, N.; Andersen, P.Ø.; Mosavi, A.; Shamshirband, S. Rigorous connectionist models to predict carbon dioxide solubility in various ionic liquids. *Appl. Sci.* **2020**, *10*, 304. [CrossRef]
32. Zeng, B.; Guo, J.; Zhang, F.; Zhu, W.; Xiao, Z.; Huang, S.; Fan, P. Prediction model for dissolved gas concentration in transformer oil based on modified grey wolf optimizer and LSSVM with grey relational analysis and empirical mode decomposition. *Energies* **2020**, *13*, 422. [CrossRef]
33. Guo, Y.; Xu, Y.-P.; Sun, M.; Xie, J. Multi-step-ahead forecast of reservoir water availability with improved quantum-based GWO coupled with the AI-based LSSVM model. *J. Hydrol.* **2021**, *597*, 125769. [CrossRef]
34. Zhang, L.; Ge, R.; Chai, J. Prediction of China's energy consumption based on robust principal component analysis and PSO-LSSVM optimized by the Tabu search algorithm. *Energies* **2019**, *12*, 196. [CrossRef]
35. Song, Y.; Xie, X.; Wang, Y.; Yang, S.; Ma, W.; Wang, P. Energy consumption prediction method based on LSSVM-PSO model for autonomous underwater gliders. *Ocean Eng.* **2021**, *230*, 108982. [CrossRef]
36. Bemani, A.; Baghban, A.; Mohammadi, A.H.; Andersen, P. Østebø Estimation of adsorption capacity of $CO_2$, $CH_4$, and their binary mixtures in Quidam shale using LSSVM: Application in $CO_2$ enhanced shale gas recovery and $CO_2$ storage. *J. Nat. Gas Sci. Eng.* **2020**, *76*, 103204. [CrossRef]
37. Yuan, X.; Chen, C.; Yuan, Y.; Huang, Y.; Tan, Q. Short-term wind power prediction based on LSSVM–GSA model. *Energy Convers. Manag.* **2015**, *101*, 393–401. [CrossRef]
38. Lu, P.; Ye, L.; Sun, B.; Zhang, C.; Zhao, Y.; Teng, J. A new hybrid prediction method of ultra-short-term wind power forecasting based on EEMD-PE and LSSVM optimized by the GSA. *Energies* **2018**, *11*, 697. [CrossRef]
39. Li, K.; Liang, C.; Lu, W.; Li, C.; Zhao, S.; Wang, B. Forecasting of Short-Term Daily Tourist Flow Based on Seasonal Clustering Method and PSO-LSSVM. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 676. [CrossRef]
40. Esene, C.; Zendehboudi, S.; Shiri, H.; Aborig, A. Deterministic tools to predict recovery performance of carbonated water injection. *J. Mol. Liq.* **2020**, *301*, 111911. [CrossRef]
41. Afzali, S.; Zendehboudi, S.; Mohammadzadeh, O.; Rezaei, N. Hybrid mathematical modelling of three-phase flow in porous media: Application to water alternating gas injection. *J. Nat. Gas Sci. Eng.* **2021**, *94*, 103966. [CrossRef]
42. Menad, N.A.; Noureddine, Z. An efficient methodology for multi objective optimization of water alternating $CO_2$ EOR process. *J. Taiwan Inst. Chem. Eng.* **2019**, *99*, 154–165. [CrossRef]
43. Amar, M.N.; Zeraibi, N.; Jahanbani Ghahfarokhi, A. Applying hybrid support vector regression and genetic algorithm to water alternating $CO_2$ gas EOR. *Greenh. Gases Sci. Technol.* **2020**, *10*, 613–630. [CrossRef]
44. Nwachukwu, A.; Jeong, H.; Sun, A.; Pyrcz, M.; Lake, L.W. Machine learning-based optimization of well locations and WAG parameters under geologic uncertainty. In Proceedings of the SPE Improved Oil Recovery Conference, Tulsa, OK, USA, 14–18 April 2018.
45. You, J.; Ampomah, W.; Sun, Q.; Kutsienyo, E.J.; Balch, R.S.; Dai, Z.; Zhang, X. Machine learning based co-optimization of carbon dioxide sequestration and oil recovery in $CO_2$-EOR project. *J. Clean. Prod.* **2020**, *260*, 120866. [CrossRef]
46. You, J.; Ampomah, W.; Sun, Q. Co-optimizing water-alternating-carbon dioxide injection projects using a machine learning assisted computational framework. *Appl. Energy* **2020**, *279*, 115695. [CrossRef]

47. Nygård, J.I.; Andersen, P.Ø. Simulation of Immiscible Water-Alternating-Gas Injection in a Stratified Reservoir: Performance Characterization Using a New Dimensionless Number. *SPE J.* **2020**, *25*, 1711–1728. [CrossRef]
48. Ripley, B.D. *Pattern Recognition and Neural Networks*; Cambridge University Press: Cambridge, UK, 2007.
49. Haykin, S. *Neural Networks and Learning Machines*, 3rd ed.; Pearson Education: Upper Saddle River, NJ, USA, 2010.
50. Fuks, O.; Tchelepi, H.A. Limitations Of Physics Informed Machine Learning For Nonlinear Two-Phase Transport In Porous Media. *J. Mach. Learn. Model. Comput.* **2020**, *1*, 19–37. [CrossRef]
51. Karniadakis, G.E.; Kevrekidis, I.G.; Lu, L.; Perdikaris, P.; Wang, S.; Yang, L. Physics informed machine learning. *Nat. Rev. Phys.* **2021**, *3*, 422–440. [CrossRef]
52. Kennedy, J.; Eberhart, R. Particle swarm optimization. In Proceedings of the IC-NN'95—International Conference on Neural Networks, Perth, WA, Australia, 27 November–1 December 1995; Volume 4, pp. 1942–1948.
53. Poli, R.; Kennedy, J.; Blackwell, T. Particle swarm optimization. *Swarm Intell.* **2007**, *1*, 33–57. [CrossRef]
54. Rashedi, E.; Nezamabadi-Pour, H.; Saryazdi, S. GSA: A gravitational search algorithm. *Inf. Sci.* **2009**, *179*, 2232–2248. [CrossRef]
55. Mirjalili, S. Genetic algorithm. In *Evolutionary Algorithms and Neural Networks*; Springer: Cham, Switzerland, 2019; pp. 43–45.
56. Mirjalili, S.; Mirjalili, S.M.; Lewis, A. Grey wolf optimizer. *Adv. Eng. Softw.* **2014**, *69*, 46–61. [CrossRef]
57. Nash, J.E.; Sutcliffe, J.V. River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol.* **1970**, *10*, 282–290. [CrossRef]
58. Moriasi, D.N.; Arnold, J.G.; Van Liew, M.W.; Bingner, R.L.; Harmel, R.D.; Veith, T.L. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE* **2007**, *50*, 885–900. [CrossRef]
59. Székely, G.J.; Rizzo, M.L.; Bakirov, N.K. Measuring and testing dependence by correlation of distances. *Ann. Stat.* **2007**, *35*, 2769–2794. [CrossRef]