



Delft University of Technology

#### Document Version

Final published version

#### Citation (APA)

Tian, X., Asadiesfahani, M., Groenenboom, J., & Voskov, D. V. (2022). Stochastic Discrete Well Affinity (DiWA) Model for Data Quality Diagnostic and Production Forecast. In *European Conference on the Mathematics of Geological Reservoirs 2022* <https://doi.org/10.3997/2214-4609.202244097>

#### Important note

To cite this publication, please use the final published version (if applicable).  
Please check the document version above.

#### Copyright

In case the licence states "Dutch Copyright Act (Article 25fa)", this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.  
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

#### Sharing and reuse

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

#### Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.  
We will remove access to the work immediately and investigate your claim.

*This work is downloaded from Delft University of Technology.*

***Green Open Access added to TU Delft Institutional Repository***

***'You share, we take care!' - Taverne project***

**<https://www.openaccess.nl/en/you-share-we-take-care>**

Otherwise as indicated in the copyright section: the publisher is the copyright holder of this work and the author uses the Dutch legislation to make this work public.

# Stochastic Discrete Well Affinity (DiWA) model for data quality diagnostic and production forecast

X. Tian<sup>1</sup>, M. Asadiesfahani<sup>1</sup>, J. Groenenboom<sup>3</sup>, D. Voskov<sup>1,2</sup>

<sup>1</sup> Delft University of Technology; <sup>2</sup> Stanford University; <sup>3</sup> Shell

## Summary

---

In this study, we present a history matching framework for oil production forecast based on synthetic and real production data developed using the stochastic Discrete Well Affinity (DiWA) model. With the increase in the complexity of the geological model and the uncertainty in the geological data, it becomes more difficult (sometimes infeasible) to conduct model inversion and production based on conventional technique. To address this problem, we proposed a stochastic DiWA model with unstructured low-resolution mesh to represent the location of wells and reservoir fluid dynamics. With this method, we can efficiently train the forward model based on production data and a stochastic ensemble of property realization. The performance of forward evaluation benefits from the Operator-Based Linearization (OBL) technique and the adjoint method for gradient calculation. Before the model training, a large ensemble size of stochastic DiWA models is generated based on the permeability statistics of the real reservoir, and those models are then filtered using the misfit between the true production data and the DiWA model response. The filtered models have the best fit with the production history of the real reservoir, while they also contain the basic geological information of the real field. The proposed method is tested first on a synthetic data ensemble for production forecast and then applied to a real field. Based on real observations, we use the DiWA model for data quality diagnostic and identify certain flaws in the collected data and model assumptions. Based on these findings, the original assumptions and data observations have been adjusted and the resulting DiWA model was successfully trained. The prediction quality of the trained DiWA model is comparable to conventional simulation techniques based on detailed geological models and has the advantage of a much more efficient and faster ability to update and maintain the subsurface model when continuous updates in production data become available. This study shows that the proposed method can provide the history matching results with high accuracy and low computational costs. Furthermore, the performance of the stochastic DiWA model can be further improved using more comprehensive and geologically constrained initial and boundary conditions.

## Introduction

With the increasing demand for efficient reservoir simulation, various methods are developed to reduce computational costs. At the same time, more and more complex and intricate geological structures and models are introduced to achieve a higher accuracy of the simulation results. However, balancing the accuracy and the efficiency is not a trivial work, as the utilization of the complex and high-resolution model may cause a rapid increase in the computational efforts. This conclusion is also valid for conducting model inversion and forecast.

Many approaches can be applied to reduce the computational cost. The upscaling method is one of the most often used approaches to reduce the degrees of freedom of the simulation problem (Durlofsky, 2005). The idea of the upscaling method is to replace the high-resolution model with the low-resolution mesh while constraining the model simulation error into an acceptable range. The multi-scale method introduces a concept of basis function, which helps to communicate between the simulation results on the coarse grid and the high-resolution grid (Jenny et al., 2003). The reservoir simulation, which is the most time-consuming part of the computation, is now performed on the coarse grid. The results on the high-resolution grid are then reconstructed based on the basis functions. Tian et al. (2021) proposed a data-driven Discrete Well Affinity (DiWA) model that is calibrated and driven by observation data. The mesh grid used in the DiWA model is very coarse, and it can be constructed only based on limited geological information.

The data-driven method is usually applied when the geological information of the reservoir is not sufficient and the amount observation data is large. It involves low degrees of freedom of the proxy model which is sufficient to represent a real reservoir response after a certain proxy model training. For example, the capacitance/resistance model (Albertoni and Lake, 2003) use relatively low degrees of freedom of the proxy model to represent the reservoir. The flow-network model decouples the full 3D flow into a set of 1D finite-difference reservoir models to reduce the computational costs (Lerlerpakdee et al., 2014). Zhao et al. (2015) proposed the interwell numerical simulation model (INSIM) and it is later extended to the interwell numerical simulation model with front-tracking (INSIM-FT) for 3D multilayer reservoirs (Guo et al., 2018).

Once the objective function of the data-driven model is constructed, it is essential to find the optimum efficiently and accurately. For the selection of optimization methods, it can be either gradient-based or gradient-free algorithms. We will only discuss the gradient-based algorithms in this study, considering that the gradient-free algorithms often have a limited speed of convergence.

If the gradients are calculated numerically in gradient-based method, it can be very time-consuming as a large amount of forward simulation runs are needed to get the partial derivatives with respect to each model control variable. Moraes et al. (2017) pointed out that it is possible to generate unreasonable gradients when perturbation of control variables is not compatible for a given optimization problem. In any case, the adjoint method is a good choice to evaluate the gradients because of its high efficiency and accuracy. The idea of adjoint method is combining the objective function and the reservoir governing equations by introducing an extra Lagrange multiplier, while this procedure does not change the stationary point of the original objective function (Jansen, 2011).

Tian et al. (2021) implemented the adjoint method within the Delft Advanced Research Terra Simulator (DARTS). The adjoint gradients have been first tested on two large ensembles of fluvial models, and then applied on proxy model for Brugge field for model training and forecast. Later, adjoint gradients allowed to develop stochastic Discrete Well Affinity (DiWA) data-driven physics-based proxy model.

In this study, we will use the stochastic DiWA framework for data quality diagnostic and production forecast. The reliability and the accuracy of the stochastic DiWA model has been shown on the synthetic Brugge model (Tian and Voskov, 2022). The application of stochastic DiWA model to a real oil field is demonstrated in this study. The model training are based on the coarse mesh grids and production data. Several training attempts are conducted for the field data quality diagnostic. After each diagnostic, the

corresponding corrections are applied to the model to improve the training results. These corrections include connection of an active aquifer to the model, adjustment of the initial water saturation, elimination of the abnormal data points, re-scaling the weights of depletion and flooding periods, and correcting of boundary conditions. With the diagnostics and the corrections, we conduct a final training that considers the depletion and flooding periods. The results indicate that stochastic DiWA model framework is able to fit the field production data and forecast the production rates with high accuracy and efficiency.

### Forward modeling formulation

In this section, we will explain more details about the Delft Advance Research Terra Simulator (DARTS) framework. The OBL technique is implemented in DARTS to facilitate and accelerate the calculations of the complex physics and the Jacobian assembly. Next, we will present the formulation of the governing equations and the operator-based linearization procedure of these governing equations.

The governing equations describing conservation of mass in discrete form utilize the finite-volume method and the backward Euler approximation for spatial and temporal approximation, respectively. The discrete form of governing equations for component  $c$  can be written as:

$$g_c = V \left[ \left( \phi \sum_j x_{cj} \rho_j s_j \right)^{n+1} - \left( \phi \sum_j x_{cj} \rho_j s_j \right)^n \right] - \Delta t \sum_{l \in L} \left( \sum_j x_{cj}^l \rho_j^l T_j^l \Delta \psi^l \right) + \Delta t \sum_j x_{cj} \rho_j q_j = 0, \quad (1)$$

where  $\phi$  is the porosity,  $x_{cj}$  is the mole fraction of component  $c$  in phase  $j$ ,  $s_j$  is the phase saturation,  $\rho_j$  is the phase molar density,  $V$  is the control volume,  $L$  is the interface between two neighboring grid blocks. In addition,  $\Delta \psi^l$  is the pressure difference of two neighboring grid blocks,  $T_j^l$  is phase transmissibility defined at the interface  $L$ ,  $q_j$  is the source/sink term of phase  $j$ . For simplicity, the capillary effect is neglected. Another assumption is that the instantaneous thermodynamic equilibrium is chosen to close the system.

To implement the operator-based linearization on Equation (1), a new notation  $\mathbf{w}$  is introduced to wrap up the physical state variables  $[p, z_c]$ . Apart from the physical state variables  $\mathbf{w}$ , the rest of variables are collected and defined as  $\boldsymbol{\xi}$  and  $\mathbf{w}$ , which are the notations wrapping up the spatial-related variables (e.g. permeability, porosity, etc.) and well-related variables (e.g. BHP controls, rate controls, etc.), respectively. Now, with the introduction of  $\mathbf{w}$ ,  $\boldsymbol{\xi}$ , and  $\mathbf{w}$ , Equation (1) can be re-written as (Voskov, 2017):

$$\mathbf{g}_c(\boldsymbol{\xi}, \mathbf{w}, \mathbf{w}) = V(\boldsymbol{\xi}) \phi_0(\boldsymbol{\xi}) (\alpha_c(\mathbf{w}) - \alpha_c(\mathbf{w}^n)) - \Delta t \sum_l \beta_c^l(\mathbf{w}) T^{ab}(\boldsymbol{\xi}) (p^b - p^a) + \theta_c(\boldsymbol{\xi}, \mathbf{w}, \mathbf{w}) = 0, \quad (2)$$

where

$$\alpha_c(\mathbf{w}) = (1 + c_r(p - p_0)) \sum_{j=1}^{n_p} x_{cj} \rho_j s_j, \quad (3)$$

$$\beta_c(\mathbf{w}) = \sum_{j=1}^{n_p} x_{cj} \frac{k_{rj}}{\mu_j} \rho_j, \quad (4)$$

$$\theta_c(\boldsymbol{\xi}, \mathbf{w}, \mathbf{w}) = \Delta t \sum_{j=1}^{n_p} x_{cj} \rho_j q_j(\boldsymbol{\xi}, \mathbf{w}, \mathbf{w}), \quad (5)$$

$c_r$  is the rock compressibility,  $\phi_0$  is the original porosity,  $p_0$  is the reference pressure,  $T^{ab}$  is the transmissibility between cell  $\mathbf{a}$  and  $\mathbf{b}$ ,  $\mathbf{w}$  and  $\mathbf{w}^n$  are physical state unknowns at the current and the previous timestep, respectively.

As it can be seen from Equation (2),  $\alpha_c$  and  $\beta_c$  are only dependent on  $\mathbf{w}$ . This indicates that when assembling the residuals and the Jacobian,  $\alpha_c$  and  $\beta_c$  can be adaptively computed at uniform grid in parameter space  $\mathbf{w}$  and approximated using multilinear interpolation (Khait and Voskov, 2018). This

technique reduces routine computation of complex nonlinear physical properties which often includes solution of various nonlinear equations (e.g., multiphase flash). The interpolation technique applied in OBL method can cause some inaccuracy but at the same time, it largely reduces the computational time required in simulation. To balance the accuracy of the physics and the performance of the simulator, an appropriate OBL space resolution should be selected. More discussions about this topic can be found in Voskov (2017).

## History matching and adjoint gradients formulation

Before doing the production forecast, a history matching of reservoir simulation should be conducted. The history matching problem aims to reduce the difference between the observation data and the model response. The first step of conducting history matching is the construction of the objective function. Generally, the well rate and BHP are the most available data for an oil field, because there are plenty of production data recorded in the course of oil field development. Therefore, the well rate and BHP data can be encapsulated into the objective function for history matching. The expression of objective function reads:

$$J(\boldsymbol{\omega}, \boldsymbol{u}, \boldsymbol{w}) = (G(\boldsymbol{\omega}, \boldsymbol{u}, \boldsymbol{w}) - \boldsymbol{d}_{\text{obs}})^T C_D^{-1} (G(\boldsymbol{\omega}, \boldsymbol{u}, \boldsymbol{w}) - \boldsymbol{d}_{\text{obs}}), \quad (6)$$

where  $J$  denotes the objective function,  $G$  is the model response,  $\boldsymbol{u}$  is the model control variables,  $\boldsymbol{d}_{\text{obs}}$  represents the observation data (i.e. well rate or BHP data), and  $C_D^{-1}$  is the inverse of the diagonal matrix about measurement error. History matching problem aims to minimize the value of the objective function by adjusting the values of model control variables  $\boldsymbol{u}$ . These model control variables are explained in the next section.

### Model control variables

In this study, three types of variables are selected to be the model control variables in history matching: transmissibility, well index, and rock-fluid interaction parameters. Their definitions relate to reservoir geological properties, well properties, and reservoir fluid properties, respectively. Note that the magnitude of different types of model control variables can be highly different. For example, the magnitude of transmissibility is usually larger than 100, while in rock-fluid interaction relations the connate water saturation and residual oil saturation are less than 0.49. This may result in a very bad optimizer's performance. It is therefore essential to scale different types of model control variables in reasonable ranges of values. There are no strict rules to define a "reasonable" value range of the model control variables. Generally, scaling these model control variables in the range of (0, 1) is suitable for the existing optimization algorithms. We will also apply this scaling strategy to transmissibility, well index, and rock-fluid interaction parameters.

Based on prior information about the permeability, the initial guess for transmissibility of interface can be evaluated as:

$$T^{ab} = c \left( \frac{\gamma_a \gamma_b}{\gamma_a + \gamma_b} \right), \quad \gamma_i = A \frac{k_i}{D_i}, \quad (7)$$

where  $c$  is the unit conversion coefficient,  $A$  is the interface area,  $D$  is the distance from the pressure node to the interface along the line connecting two pressure nodes,  $k$  is the permeability, and the subscript  $i$  denotes the cell index. The transmissibility control variables can therefore be initialized based on Equation (7). Moreover, the gradients of objective function with respect to transmissibility control variables will be calculated by using adjoint method, which highly improves the calculation efficiency with two to three orders of magnitude improvement in comparison to numerical gradients.

Similar relationships can be applied for well index evaluation using modified Peaceman formula (Peaceman, 1983):

$$T^w = c \frac{2\pi\Delta z \sqrt{k_x k_y}}{\ln r_o/r_w + S}. \quad (8)$$

Apparently, Equation (8) is not applicable in unstructured grid for calculating the well index. However, this equation can still be utilized to provide initial guess for well indexes. The gradients of objective

function with respect to well index will also be prepared by the adjoint method.

Finally, rock-fluid interaction control variables should define fluid properties using Brooks-Corey correlations:

$$k_{ro} = k_{ro}^e (1 - S_w^*)^{n_o}, \quad (9)$$

$$k_{rw} = k_{rw}^e (S_w^*)^{n_w}, \quad (10)$$

$$S_w^* = \frac{S_w - S_{wc}}{1 - S_{wc} - S_{or}}. \quad (11)$$

where  $S_w^*$  is the normalized water saturation;  $k_{rw}$  is the water relative permeability;  $k_{rw}^e$  and  $k_{ro}^e$  are the endpoint relative permeability of water and oil, respectively;  $n_w$  and  $n_o$  are the exponents of water and oil, respectively;  $S_w$  is the water saturation;  $S_{wc}$  is the connate water saturation;  $S_{or}$  is the residual oil saturation. To control physical properties in the governing equations, we only need the following six variables as the rock-fluid interaction control variables:

$$\{S_{or}, S_{wc}, n_o, n_w, k_{rw}^e \frac{\rho_w}{\mu_w}, k_{ro}^e \frac{\rho_o}{\mu_o}\}. \quad (12)$$

For simplicity, the gradients of objective function with respect to these six control variables are calculated numerically.

### Adjoint gradients

In the previous section, we have introduced transmissibility, well index, and rock-fluid interaction variables used in history matching. To find the minimum of Equation (6), we can perform the first order derivatives of Equation (6) with respect to the aforementioned model control variables and use the gradient descent method. The first order derivatives are obtained either numerically or analytically. It is basically infeasible to calculate the numerical gradients for a very large dimensions of control space (e.g. transmissibility control variables). To solve this problem, we implemented the adjoint method in DARTS framework to compute the gradients analytically and efficiently. The next derivations will mostly follow the notation from Jansen (2011) which were adapted to OBL framework in Tian et al. (2021).

In the adjoint method, instead of directly performing the first order derivatives of the objective function, we first combine the objective function with the reservoir governing equations and then calculate the derivatives. The new equation reads:

$$\tilde{\mathcal{J}}(\boldsymbol{\omega}, \mathbf{u}, \boldsymbol{\lambda}) = J(\boldsymbol{\omega}, \mathbf{u}) + \boldsymbol{\lambda}^T g(\boldsymbol{\omega}, \mathbf{u}). \quad (13)$$

Here Equation (13) is called augmented objective function. Note that in this equation we omit the notation of well control variable  $\mathbf{w}$ . The model control variables  $\mathbf{u}$  actually corresponds to the spatial-related variables  $\boldsymbol{\xi}$  in Equation (2).  $\boldsymbol{\lambda}^T$  is the transposed Lagrange multiplier.

To find the optimum of the augmented objective function, the following equation should be satisfied:

$$\tilde{\mathcal{J}}_{\boldsymbol{\lambda}} = g(\boldsymbol{\omega}, \mathbf{u}) = 0, \quad (14)$$

$$\tilde{\mathcal{J}}_{\boldsymbol{\omega}} = \boldsymbol{\lambda}^T g_{\boldsymbol{\omega}}(\boldsymbol{\omega}, \mathbf{u}) + J_{\boldsymbol{\omega}}(\boldsymbol{\omega}, \mathbf{u}) = 0, \quad (15)$$

$$\tilde{\mathcal{J}}_{\mathbf{u}} = \boldsymbol{\lambda}^T g_{\mathbf{u}}(\boldsymbol{\omega}, \mathbf{u}) + J_{\mathbf{u}}(\boldsymbol{\omega}, \mathbf{u}) = 0, \quad (16)$$

where the subscript  $\boldsymbol{\lambda}$ ,  $\boldsymbol{\omega}$  and  $\mathbf{u}$  represents the derivatives with respect to corresponding variables. Obviously, Equation (14) is already satisfied because it is actually Equation (2) and solved in forward simulation. The newly introduced  $\boldsymbol{\lambda}^T$  is solved from Equation (15). Eventually,  $\tilde{\mathcal{J}}_{\mathbf{u}}$  is assembled and combined with the numerical gradients of rock-fluid variables for gradient descent. More details and discussions about the adjoint method can be found in Jansen (2011) and Tian and Voskov (2022).

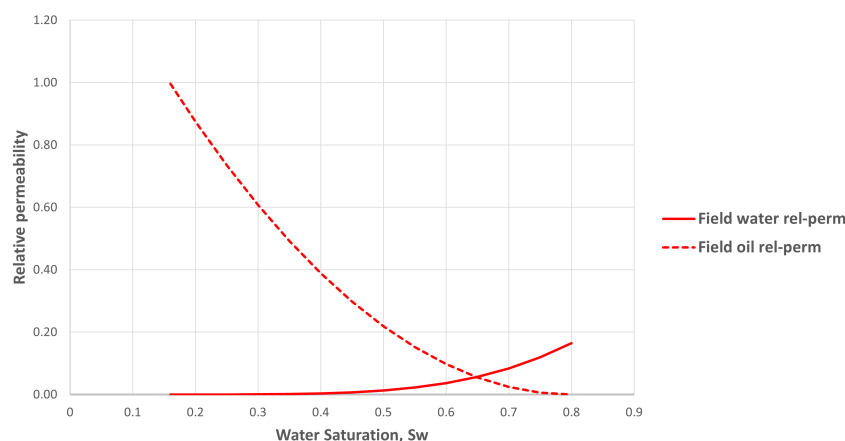
## Results

An application of stochastic DiWA model to a synthetic reservoir field was described in Tian and Voskov (2022). In this section, we present an application of stochastic DiWA methodology to a sector of a real hydrocarbon reservoir.

### *Initial model construction*

We will start with reservoir description and DiWA model construction. Later, data quality diagnostic and production forecast on this oil field are presented and discussed. The reservoir consists of two payzones that are not connected with each other because of a low permeable zone in between. An aquifer is connected to the lower part of the payzone. This aquifer acts as the water drive mechanism for oil production. A major normal fault (called Central-HC fault) exists in this reservoir along the North-East to South-West direction. It is assumed that there is no water flowing through this fault. Apart from Central-HC fault, there are some small permeable faults in the reservoir. The reservoir is located in a rectangular zone of 1500 by 500 m. The top surface of the reservoir is at the depth of 640 m. The net thickness is 103 m. The water-oil contact is at the depth of 653 m. In total, 43 producers and 18 injectors are placed in the reservoir. Both the horizontal wells and the vertical wells are applied in the payzone. There are 16 horizontal wells (including 13 producers and 3 injectors) with open-hole completion in total. The rest are all vertical wells with cased-hole completion. Field production data are recorded and collected at the surface condition from the separators.

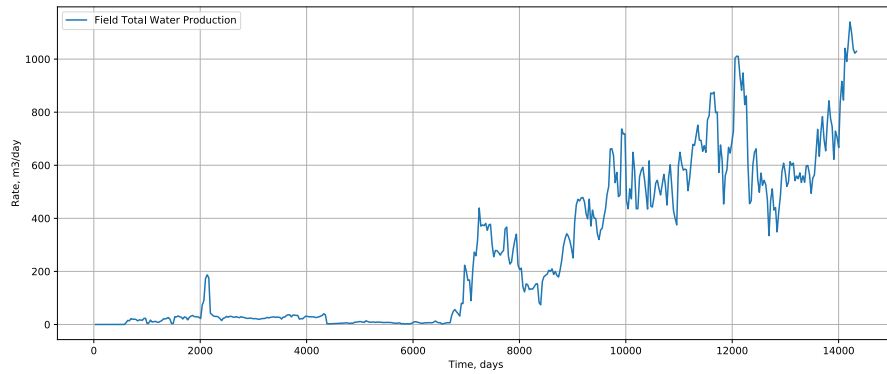
The petrophysical data show that the aquifer has the highest permeability, while the upper payzone has the moderate permeability and the lower payzone is the least permeable. The lower payzone has higher initial water saturation because it connects directly to the aquifer. The relative permeability curves are based on core test analysis, see Figure 1.



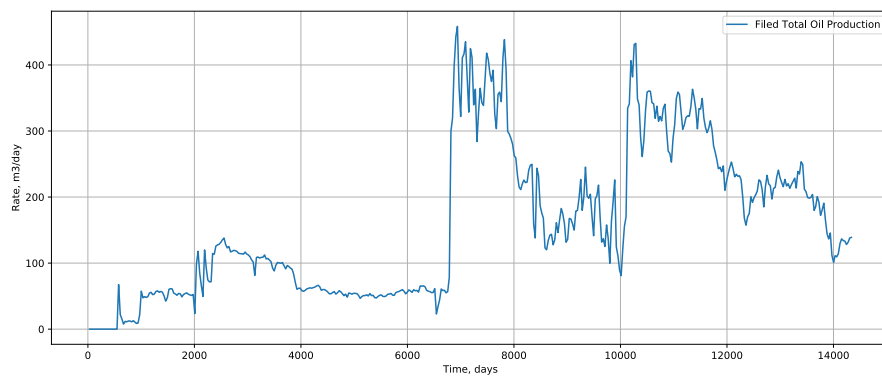
**Figure 1** The relative permeability curves of oil and water

The viscosity of oil is 93 cP at the condition of 94 bar and 45 °C. The primary depletion period ends at around 7000 days, as it can be seen from Figure 2 that there is a water production jump at this time. After the water breakthrough, the injectors start to inject water into the reservoir. It can be noticed that after 7000 days, the oil production is higher than before because of the water flooding, see Figure 3.

Apart from the aforementioned basic information of the reservoir, a physical model is needed and discretized spatially. In this study, we only consider sector of the entire model. We build a coarse unstructured grid mesh that is constrained by the sector contour. The mesh contains two layers, as there are two payzones that are disconnected in the reservoir. Now, we can assign the initial guess for the properties based on existing high-fidelity model. In Figure 4, the porosity distribution from high-fidelity reservoir model is projected to the DiWA model as the initial porosity map by using the volume-weighted

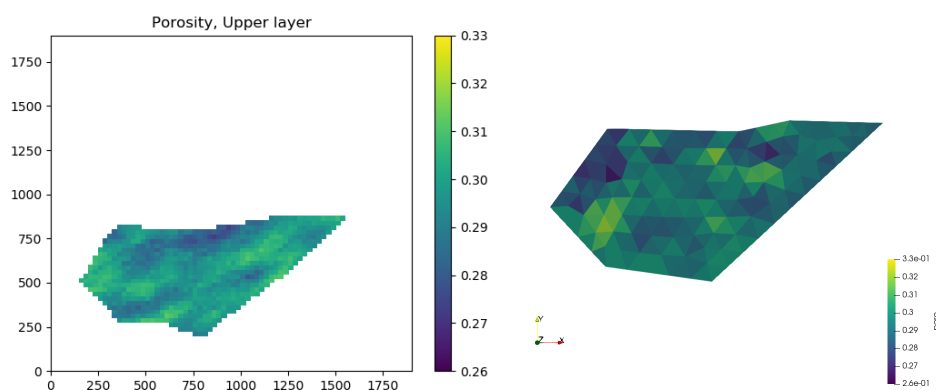


**Figure 2** The field total water production curve



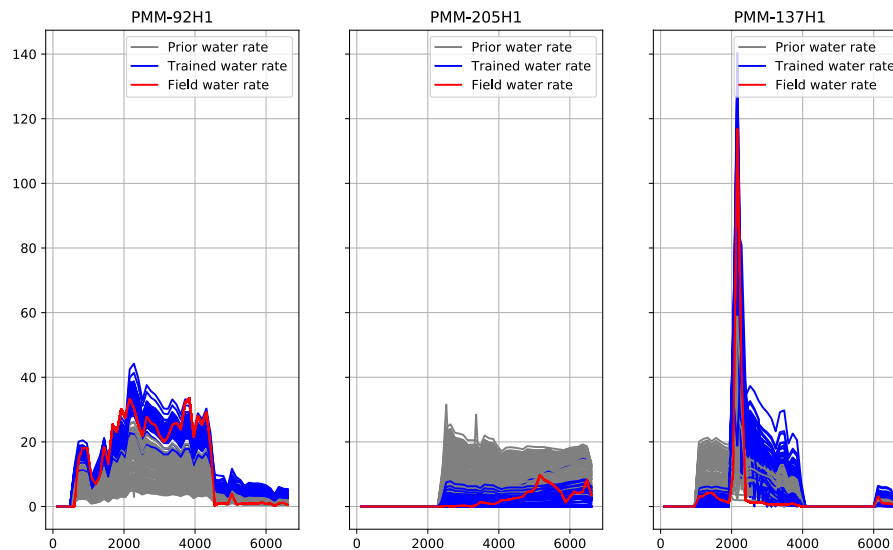
**Figure 3** The field total oil production curve

geometrical average method. The simulation uses dead oil model. The initial conditions of DiWA model, including initial reservoir pressure, well locations, average water saturation, and average depth, are consistent with the high-fidelity model. This is done for simplicity and can be easily replaced by approximate data as in Tian and Voskov (2022).



**Figure 4** The porosity distribution of reservoir. The left figure shows the original porosity distribution. The right figure shows the initial porosity map in DiWA model

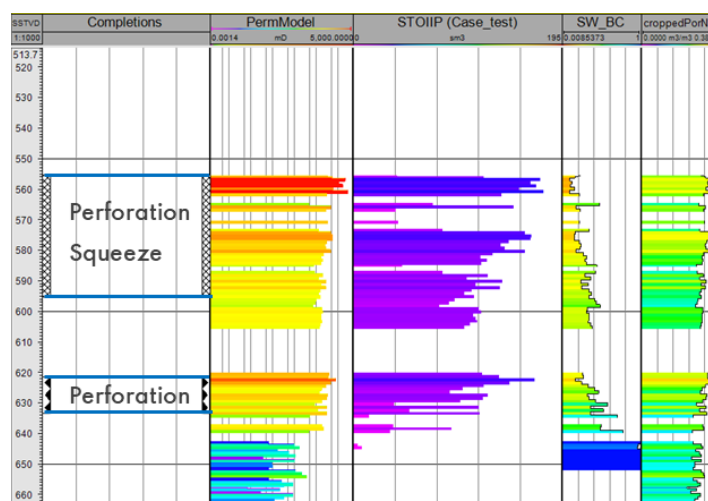
Since the magnitude of the fluid rates in the depletion period is very different from the flooding period, we decided to start our training within the depletion period. The training results of the depletion period are shown in Figure 5.



**Figure 5** The training results of water rates for the well PMM-92H1, PMM-205H1, and PMM-137H1 in the depletion period. The gray and blue curves are the water rates before (priors) and after training, respectively. The red curves are the field water rate (true data)

#### Data quality diagnostics for the depletion period

In Figure 5, we selected and showed three typical problematic wells that demonstrate obvious deviation from historic rates even at prior models generation. For example, it can be noticed that there is an abnormal water rate peak in PMM-137H1. After further investigation we learned that PMM-137H1 was completed in two payzones, see Figure 6. Later, the perforation at the upper zone was closed by a squeeze cementing operation at time 1940 days, and only the lower zone continued to produce formation liquid. This work-over operation in the field leads to a significant water peak which have no relation to reservoir dynamic but causes severe problems in the model training.



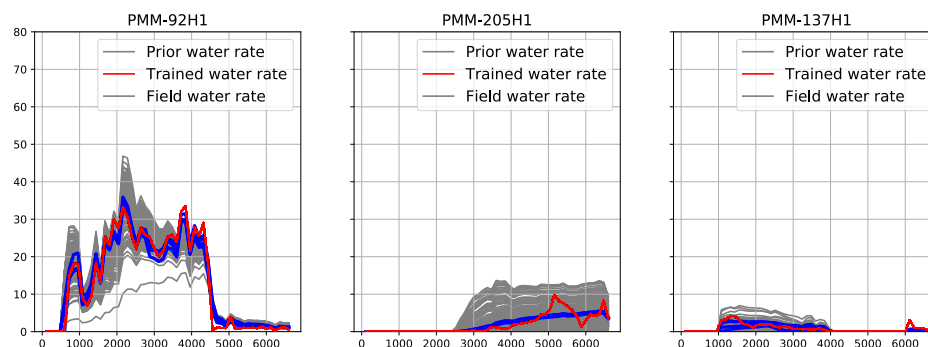
**Figure 6** The well logging interpretation results of PMM-137H1. Two perforations are located in the high permeable zones (red and orange color lines at the second column)

For PMM-92H1, most priors (the gray curves in Figure 5) have the water rates that are lower than the true field data. This is an indication that the initial assumptions may be biased from the reasonable model parameter range. Based on the field information, PMM-92H1 was completed in both upper and

lower payzones. There is an active aquifer attached to the lower payzone. It is reasonable to deduce that the high water rate of PMM-92H1 shown in the true field data in the depletion period is from the active aquifer. Similarly, the overestimated priors compared with the true field data of PMM-205H1 may indicate the biased initial guesses of the model.

Based on the previous analysis and the basic geological information, it can be noticed that there is a clear border between these two payzones. The well productions are largely affected by the different properties and field operations implemented in these two payzones. Therefore, it is necessary to apply a correction that the rock-fluid interaction control variables are trained separately in two different payzones. The correction corresponding to PMM-92H1 is attaching an aquifer to the lower payzone that supplies the extra water production shown in PMM-92H1 of Figure 5. Considering the priors of PMM-205H1, the correction of the initial saturation distribution of DiWA model is applied by interpolating based on the hard data at the well. As for the abnormal water rate peak in PMM-137H1, the datum at that peak is removed from the true field data before training.

With the data diagnostic and the corrections applied on the initial guesses, true field data, and training strategies, the results show that the priors follow the trend of the field data, and the trained results are more accurate, see Figure 7. The training results of the total rates of all activated wells can be found in Figure 8.



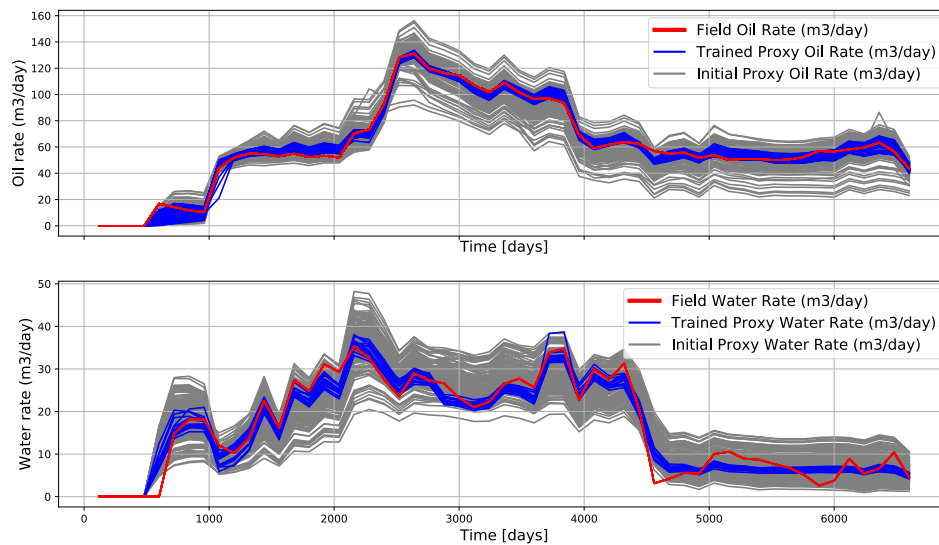
**Figure 7** The training results of water rates for the well PMM-92H1, PMM-205H1, and PMM-137H1 in the depletion period after the data diagnostic and the corrections. The gray and blue curves are the water rates before (priors) and after training, respectively. The red curves are the field water rates (true data)

### Full model training and production forecast

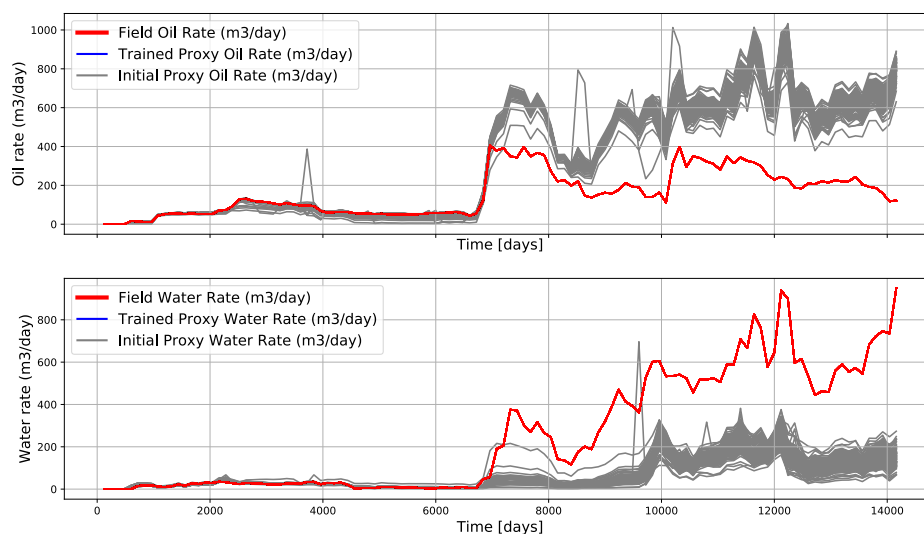
To include both the depletion and flooding periods, the second attempt of training was conducted. The priors can be directly inherited from the trained DiWA model of the depletion period. However, Figure 9 shows that these priors can not adequately represent the trend of the flooding period. The trained DiWA model in the depletion period makes the water less mobile than the oil in the flooding period. This is because the driven force and mechanism of the depletion period are very different from the flooding period. In this case, the priors need to be updated and filtered to generate new priors that follow the trend of both depletion and flooding periods.

Note that there are big differences in the magnitude of rates between the depletion and flooding periods, which will cause disproportionate weights in the objective function. We re-scale the weights of these two periods in both the filtering and training procedures. Figure 10 demonstrates the training and forecast results under the new weighting strategy and the update of priors.

As it can be seen from Figure 10, the trained DiWA model fit the training period data (light green area) very well for both the oil and water rates, while the curves at the forecast period (light red area) deviate from the field total data to some extent. To solve this problem, we need to review the field information



**Figure 8** The total oil and water rates of all activated wells in the depletion period. The gray and blue curves are the total rates before (priors) and after training, respectively. The red curves are the field total rates (true data)

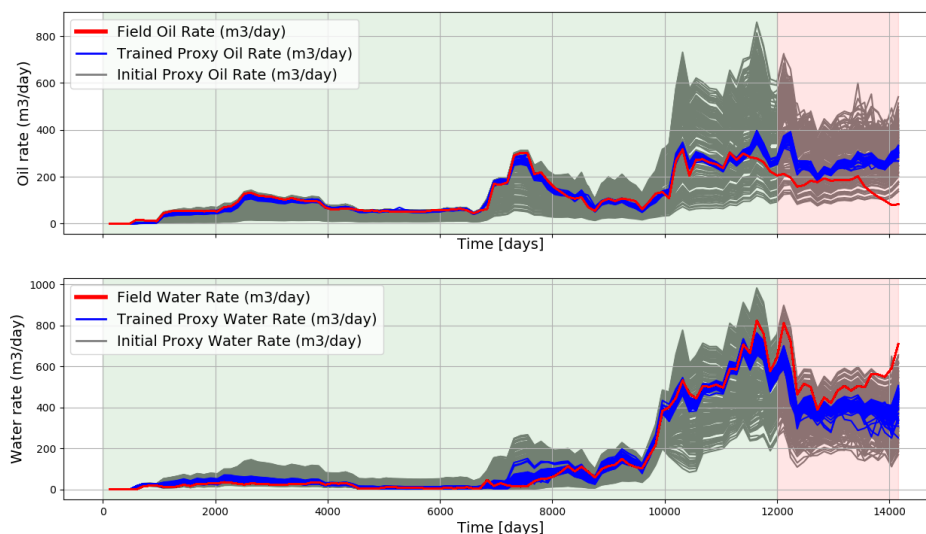


**Figure 9** The total oil and water rates of both depletion and flooding periods based on the trained DiWA model in the depletion period. The gray and red curves are the priors and field data, respectively

about wells and reservoir. Figure 11 shows the details of all production wells. It is clear to see that the trained curves of some wells deviate from the field data from 10000 days to 12000 days (e.g. PMM-92H1 and PMM-609H1). The trained curves of many wells do not converge to the field data, though they are able to follow the trend of the field data (e.g. PMM-655H1, PMM-552H2, etc.).

Combined with the field information of the well locations, it is noticed that most problematic wells are the horizontal wells. Those horizontal wells are very close to the model boundary, and most of their perforations are not included in the target area but outside of the model. In this case, the original assumption of sealing faults at the boundary should be revised to consider the flows from the boundary. It is also worthy to verify whether these problematic wells actually cause the low quality of model training results.

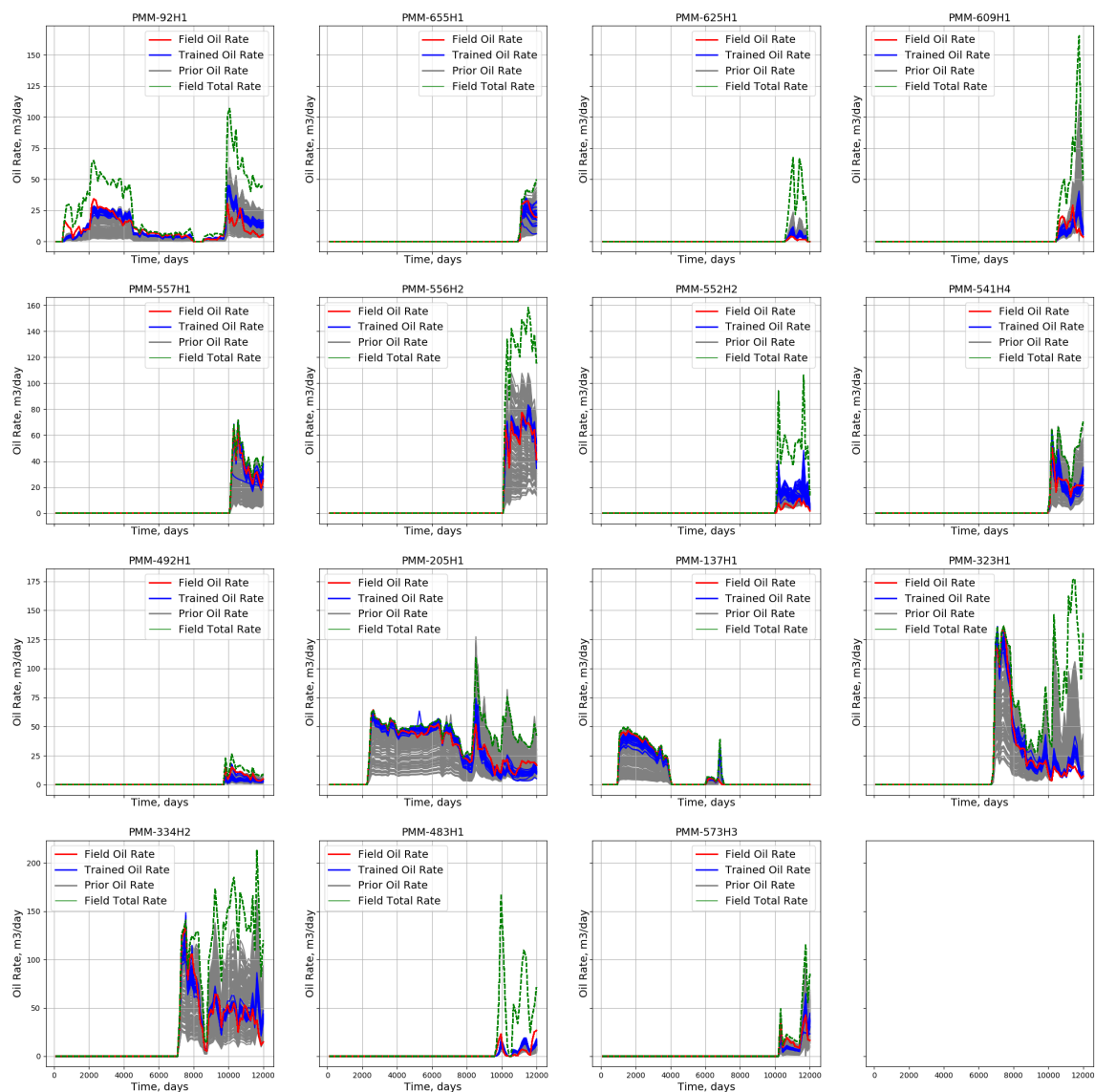
In the next step, five interior wells that are far from the model boundary, therefore, are selected to



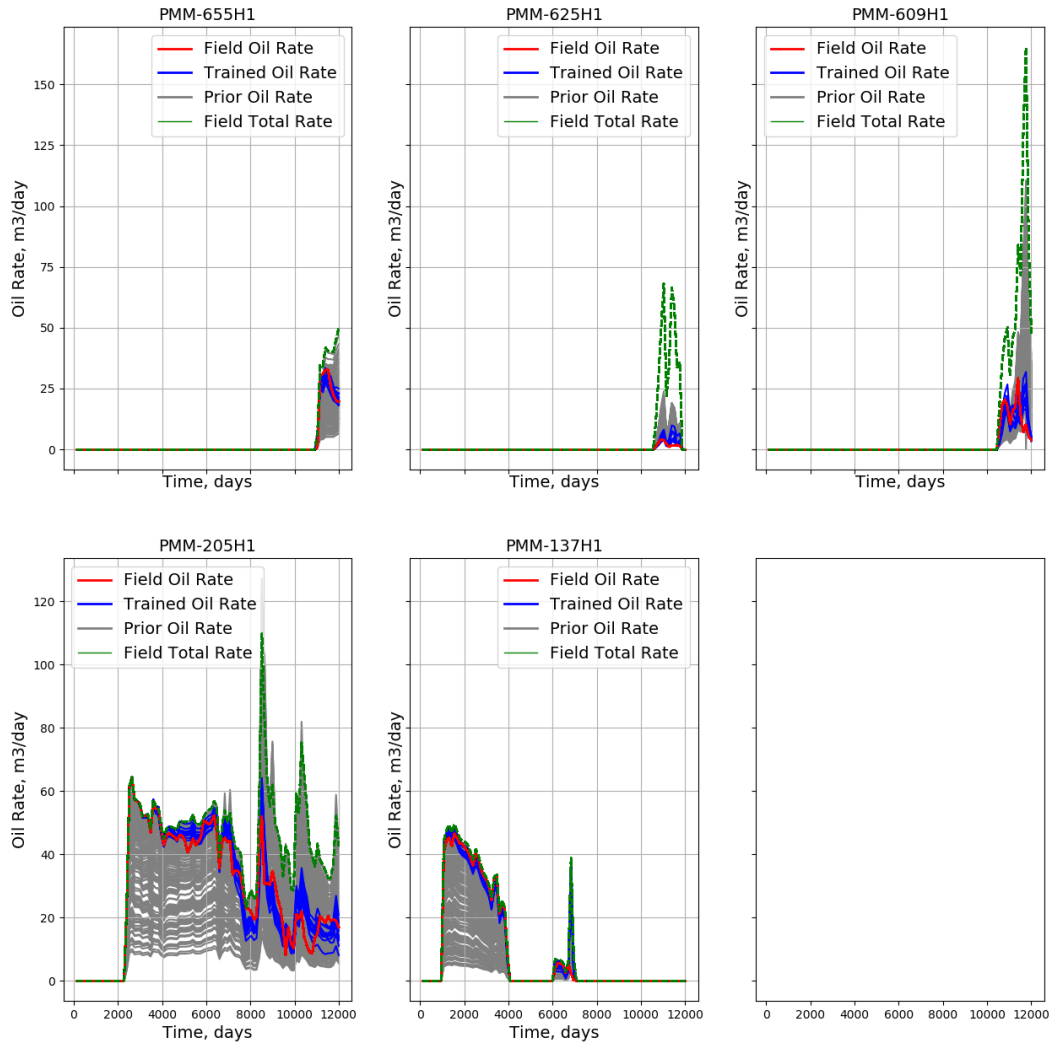
**Figure 10** The total oil and water rates for the entire production period based on the updated priors and weighting strategies. The light green and red areas represent the training and forecast periods, respectively

do a training test again, where the rest wells are excluded from the objective function but kept in the model. Figure 12 demonstrates the training results of the interior well PMM-655H1, PMM-625H1, PMM-609H1, PMM-205H1, and PMM-137H1. It can be seen that the trained results (blue curves) of PMM-655H1 and PMM-137H1 converge to the field data (red curves). The well PMM-625H1, PMM-609H1, and PMM-205H1 follow the trend of the field data. These results confirm that the faults are not sealed. Therefore, the flow from the boundary should not be neglected.

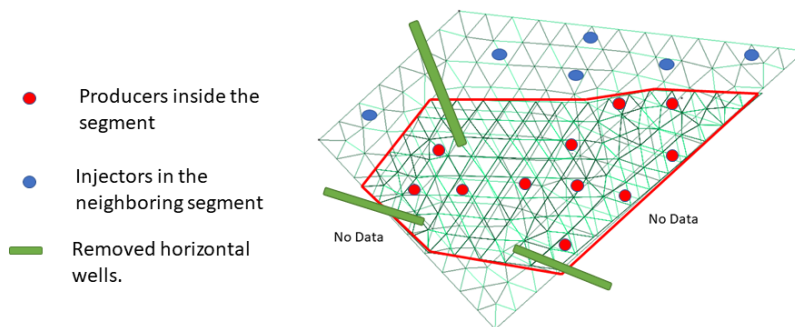
To take into account the effect of the neighboring segment outside of the boundary, the mesh grid of the DiWA model is extended. The modified mesh grid is shown in Figure 13. For those horizontal wells (PMM-323H1, PMM-334H2, PMM-573H3) that are partially outside of the model with no available data, they are removed from the training, see the green bars in Figure 13.



**Figure 11** The oil and liquid rate (oil plus water rate) of all production wells in the training period (12000 days). The gray and blue curves are the oil rates before and after training, respectively. The red and green curves are the field oil rates and liquid rates



**Figure 12** The oil and liquid rate (oil plus water rate) of interior wells (PMM-655H1, PMM-625H1, PMM-609H1, PMM-205H1, and PMM-137H1) in the training period (12000 days). The gray and blue curves are the oil rates before and after training, respectively. The red and green cures are the field oil rates and liquid rates

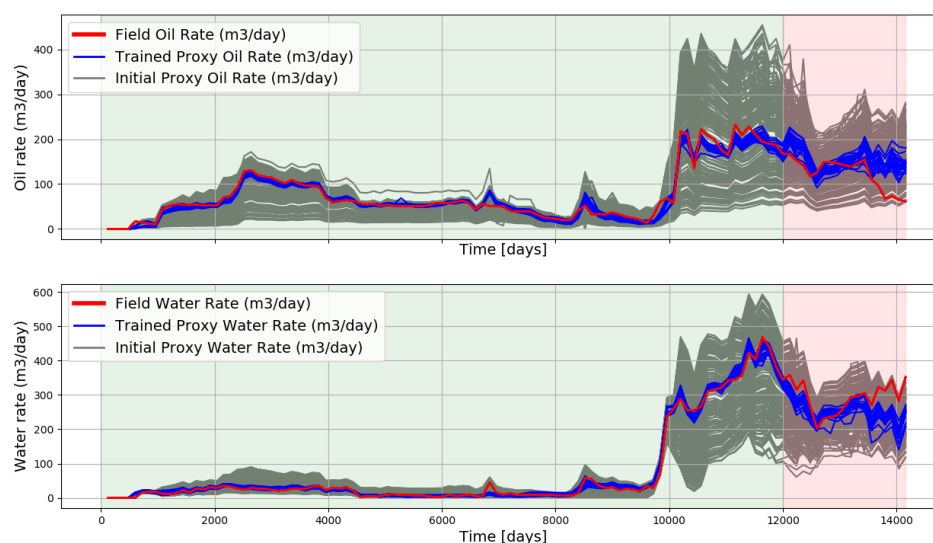


**Figure 13** The extended mesh grid of DiWA model. The original DiWA mesh grid is constrained inside the area of red lines. The extended mesh grid includes most of the boundary of the original mesh grid, while there are still some region with no available data. The red dots are the production wells to be included in the objective function. The blue dots are the approximate position of the injectors as the input of the model. The green bars represent the horizontal wells to be removed from the training.

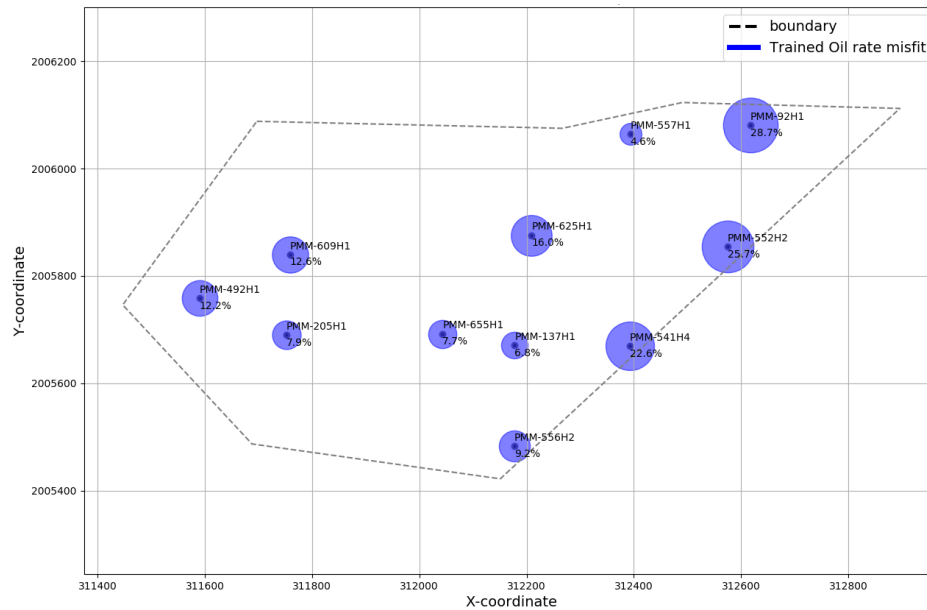
With the mesh grid extension and the removal of the horizontal wells with no available data, the final training of the DiWA model is conducted. The training and forecast results are shown in Figure 14. Compared with the previous training attempt results shown in Figure 10, the forecast period (the light red area from 12000 days to 14160 days) of Figure 14 shows a better prediction of the oil and water rates.

However, with the increase of the forecast time, the misfit between the trained DiWA model and the field data increases. This may be caused by some unexpected in-situ operations applied to the wells in the forecast period, or the significant changes of the connectivity between wells. To figure out what happened in the forecast period, we plot the oil misfit error of each well with the reservoir model, see Figure 15. PMM-552H2 and PMM-541H4 have high misfit errors. This is because they are very close to the boundary even in the extended mesh grid of DiWA model. At the same time, there is also a high misfit error for PMM-92H1, which is not expected as the boundary effect of this well is removed from the extended mesh grid.

By reviewing the field information of the well operations, it is noticed that PMM-92H1 changed from producer to injector at 13160 days, which is outside of the training period. The switching from producer to injector of PMM-92H1 may severely change the conditions of the well and its surroundings. Notice, that DiWA model is still a data-driven approach which can only capture trends incorporated into the training data. Nevertheless, Figure 14 provides an acceptable training result of the history matching and forecast of the real oil field.



**Figure 14** The final training results of the total oil and water rates. The light green and red areas represent the training and forecast periods, respectively



**Figure 15** The misfit error of oil rates for vertical wells and their well locations

## Conclusions

In this study, we present the application of the stochastic DiWA model framework on a real oil field sector. Stochastic DiWA model framework is a data-driven modeling tool that is capable of conducting history matching and forecast with high efficiency and flexibility. The high efficiency is achieved by using adjoint gradients and low-resolution mesh approximating affinity of wells. With these capabilities, we can easily conduct multiple data quality diagnostics and make the corresponding corrections by running massive amount of training attempts. These corrections include attaching the active aquifer, correcting the initial water saturation, removing the abnormal peak of water rate data, re-scaling the weights of depletion and flooding periods, and model boundary extension. These corrections to the model involve multiple re-initializations, reconstruction of the objective function, and the modification of the proxy configuration. The various corrections applied to the DiWA model also show the flexibility of the developed data-driven proxy framework. If more comprehensive and geologically constrained initial and boundary conditions are available, the performance of the stochastic DiWA model can be further improved.

## Acknowledgements

We thank the China Scholarship Council (CSC) for the financial support of X. Tian.

## References

- Albertoni, A. and Lake, L. [2003] Inferring interwell connectivity only from well-rate fluctuations in waterfloods. *SPE Reservoir Evaluation and Engineering*, **6**(1), 6–15.
- Cardoso, M., Durlofsky, L. and Sarma, P. [2009] Development and application of reduced-order modeling procedures for subsurface flow simulation. *International Journal for Numerical Methods in Engineering*, **77**(9), 1322–1350.
- Collins, D., Nghiem, L., Li, Y.K. and Grabonstotter, J. [1992] An Efficient Approach to Adaptive-Implicit Compositional Simulation With an Equation of State. *SPE Reservoir Engineering*, **7**(02), 259–264.
- Durlofsky, L. [2005] Upscaling and Gridding of Fine Scale Geological Models for Flow Simulation. *the 8th International Forum on Reservoir Simulation Iles Borromees, Stresa, Italy*.
- Guo, Z., Reynolds, A.C. and Zhao, H. [2018] A Physics-Based Data-Driven Model for History Match-

- ing, Prediction, and Characterization of Waterflooding Performance. *SPE Journal*, **23**(02), 367–395.
- Jansen, J. [2011] Adjoint-based optimization of multi-phase flow through porous media - A review. *Computers and Fluids*, **46**(1), 40–51.
- Jenny, P., Lee, S. and Tchelepi, H. [2003] Multi-scale finite-volume method for elliptic problems in subsurface flow simulation. *Journal of Computational Physics*, **187**(1), 47–67.
- Karimi-Fard, M., Durlofsky, L. and Aziz, K. [2004] An Efficient Discrete-Fracture Model Applicable for General-Purpose Reservoir Simulators. *SPE Journal*, **9**(02), 227–236.
- Khait, M. and Voskov, D. [2018] Operator-based linearization for efficient modeling of geothermal processes. *Geothermics*, **74**, 7–18.
- Lerlertpakdee, P., Jafarpour, B. and Gildin, E. [2014] Efficient production optimization with flow-network models. *SPE Journal*, **19**(6), 1083–1095.
- Lim, K.T. [1995] A New Approach for Residual and Jacobian Arrays Construction in Reservoir Simulators. *SPE Computer Applications*, **7**(04), 93–96.
- Moraes, R., Rodrigues, J., Hajibeygi, H. and Jansen, J. [2017] Multiscale gradient computation for flow in heterogeneous porous media. *Journal of Computational Physics*, **336**, 644–663.
- Peaceman, D. [1983] Interpretation of Well-Block Pressures in Numerical Reservoir Simulation With Nonsquare Grid Blocks and Anisotropic Permeability. *Society of Petroleum Engineers Journal*, **23**(03), 531–543.
- Tian, X., Blinovs, A., Khait, M. and Voskov, D. [2021] Discrete Well Affinity (DiWA) Data-Driven Proxy Model for Production Forecast. *SPE Journal*, 1–17.
- Tian, X. and Voskov, D. [2022] Efficient application of stochastic Discrete Well Affinity (DiWA) proxy model with adjoint gradients for production forecast. *Journal of Petroleum Science and Engineering*, **210**, 109911.
- Voskov, D.V. [2017] Operator-based linearization approach for modeling of multiphase multi-component flow in porous media. *Journal of Computational Physics*, **337**, 275–288.
- Zhao, H., Kang, Z., Zhang, X., Sun, H., Cao, L. and Reynolds, A. [2015] INSIM: A data-driven model for history matching and prediction for waterflooding monitoring and management with a field application. *Society of Petroleum Engineers - SPE Reservoir Simulation Symposium 2015*, **1**, 431–461.