TU Delft

Delft University of Technology

Efficient application of stochastic Discrete Well Affinity (DiWA) proxy model with adjoint gradients for production forecast

Tian, Xiaoming; Voskov, Denis

DOI
10.1016/j.petrol.2021.109911

Publication date
2022

Document Version
Final published version

Published in
Journal of Petroleum Science and Engineering

Citation (APA)
Tian, X., & 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*, Article 109911. https://doi.org/10.1016/j.petrol.2021.109911

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

Copyright
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.

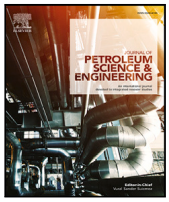
This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Journal of Petroleum Science and Engineering 210 (2022) 109911

Contents lists available at ScienceDirect

Journal of Petroleum Science and Engineering

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



Efficient application of stochastic Discrete Well Affinity (DiWA) proxy model with adjoint gradients for production forecast

Xiaoming Tian [a], Denis Voskov [a,b,*]

[a] Department of Geoscience and Engineering, TU Delft, Delft, Netherlands
[b] Department of Energy Resources Engineering, Stanford University, CA, USA

ARTICLE INFO

Keywords:
DiWA model
Adjoint gradient
Data-driven model
Proxy model
DARTS

ABSTRACT

In this paper, we describe adjoint gradient formulation for the Operator-Based Linearization modeling approach. Adjoint gradients are implemented in Delft Advanced Research Terra Simulator (DARTS) framework and applied for history matching using a proxy methodology. Due to the application of adjoint gradients, the computational efficiency of the discrete well affinity (DiWA) proxy model for production forecast is significantly improved. That allows us to derive several important extensions. The proxy methodology is further extended and validated for 3D three-phase black-oil problems. The results show that the gradient-based regression can provide good history matching and reconstruct a true petrophysical characterization when the initial guess is generated based on highly reliable geological information. For cases with a limited or not sufficient geological characterization, an efficient stochastic application of DiWA proxy model is proposed. This approach consists of massive sampling procedures for collecting different realizations based on high-fidelity statistics with filtering. These realizations are generated stochastically because they are not conditioned to any production information but the basic geological statistics of the reservoir. The trained DiWA proxy model demonstrates a small deviation between the model response and the observation data. When applying the refined DiWA model for the training, the error between the model response and observation data can be further reduced. The forecast based on the trained model has slightly larger variability but the deviation is still reasonable. The enhanced DiWA methodology presents an efficient and robust technique for creating an ensemble of stochastic proxy models that can be used in production forecast, flow diagnostic, and optimization.

1. Introduction

Production optimization has always been one of the research focuses in the field of reservoir engineering. Prior to the production optimization, the geological model should be calibrated or history matched to the production data. This history matching procedure is usually quite time-consuming because of the large amounts of gradient calculation and many iteration steps of searching for the optimal solution. A reliable reservoir numerical model is the prerequisite for doing reservoir management and decision-making in subsurface energy projects. Generally, a reservoir model with higher resolution may be seen as more realistic with detailed geological information so that the calculation output can be considered more plausible. However, the degrees of freedom in the high-resolution reservoir model may contain several thousand or millions. In this case, the computational time of simulation model can take hours or even days. Moreover, the history matching process may require a large number of forward simulation runs to obtain accurate results. This consequentially results in high computational cost of forward simulation in the course of history matching.

Simplified full-field models can be applied to increase computational efficiency. This kind of approach includes (but not limited to) upscaling, multi-scale method, and streamline simulation. The upscaling method uses a coarser grid model to mimic a high-resolution reservoir model (Durlofsky, 2005). The multi-scale method takes the fine-scale grid and locally transfers it in the global representation using the basis function concept (Jenny et al., 2003). The simulation is performed at a coarse grid with a fine grid solution (e.g. pressure) reconstructed based on pre-processed basis functions. A streamline method is an Eulerian–Lagrangian approach that translates a real fluid transport problem into a one-dimensional problem solved along streamlines (Batycky et al., 1997). The performance of these methods largely depends on the reliability of the geological characterization of the reservoir. However, in many cases, reliable geological information is questionable or may not be available.

* Corresponding author at: Department of Geoscience and Engineering, TU Delft, Delft, Netherlands.
E-mail address: D.V.Voskov@tudelft.nl (D. Voskov).

https://doi.org/10.1016/j.petrol.2021.109911
Received 8 September 2021; Received in revised form 17 November 2021; Accepted 22 November 2021
Available online 18 December 2021
0920-4105/© 2021 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

In this case, another alternative called the data-driven model can help in modeling without requiring much geological information. The data-driven model often uses a proxy representation with fewer degrees of freedom compared with the high-fidelity model. It is assumed that the degrees of freedom of the proxy model are sufficient to mimic a realistic model response after a certain period of training. Once the proxy model is trained to meet the required criteria, it can be applied to do reservoir simulations and predictions with much less computational effort because of the lower degrees of freedom. Many data-driven models were developed by different researchers. For example, Guo et al. (2018) proposed a data-driven Interwell Numerical Simulation Model with Front-Tracking (INSIM-FT). Jansen and Kelkar (1997) proposed a statistical data-driven model. In the study of Zubarev (2009), the applied data-driven models utilize artificial intelligence techniques in the course of model training.

Before training the proxy model, it is essential to construct an appropriate loss function and an efficient minimization strategy. Based on the practical use of gradient information in the course of model training, they can be classified into gradient-free and gradient-based methods. Gradient-free methods are suitable for the highly nonlinear problem with many local minima, for example, simulated annealing (Kirkpatrick et al., 1983), genetic algorithm (Holland, 1984), particle swarm algorithm (Eberhart and Kennedy, 1995), etc. For the gradient-based method, the idea is searching for the reduction of the loss function in the direction of gradient (Jansen, 2011). It is essential to efficiently calculate the Hessian using the derivatives of the loss function. A straightforward approach for obtaining gradients is to calculate its derivatives numerically. However, this procedure usually requires large amounts of computational efforts considering the high degrees of freedom of the model, and possibly generates unreasonable gradient values if the selected perturbation to the control variables is not suitable for the given nonlinearity of the problem (Moraes et al., 2017).

A promising method to resolve these issues is applying the adjoint-based technique to calculate gradients, especially for the case of high-dimensional space of control variables. This approach is based on an adjoint formulation, which combines the original loss function with other constraints following the optimal control theory. The adjoint method largely reduces the computational effort, compared with the conventional numerical gradients approach for calculating partial derivatives. It has been originally developed for the optimization of dynamical systems, for example, the flight paths of rockets and satellites (Bryson and Ho, 2018). In the field of reservoir engineering, as the development of numerical technique, adjoint-based optimization was applied in numerical reservoir simulation to perform history matching for petroleum recovery process (Mehos and Ramirez, 1989; Fathi and Ramirez, 1984; Ramirez et al., 1984), thermal recovery process (Wei et al., 1993) and so on. Later, as the conceptions of "smart well" and "smart field" arose, adjoint-based optimization was widely studied and applied in the field of reservoir engineering and reservoir management (Brouwer and Jansen, 2004; Sarma et al., 2005, 2006; Kourounis et al., 2010; Volkov and Voskov, 2016).

In this study, the DARTS framework is used to implement a data-driven Discrete Well Affinity (DiWA) model. It is following up on our recent work (Tian et al., 2021) where the DiWA model was introduced and tested for a large ensemble of fluvial proxy models as well as a real petroleum field. However, a fixed uniform permeability of 1000 mD is used to generate the initial guess of the Brugge field in the study of Tian et al. (2021). In this study, the permeability of each cell of the stochastic proxy model is randomly sampled based on the probability density of the field. Before the training of these models, a filtering procedure is applied to filter out the outliers while keeping the most promising candidates under a specific threshold. The collected candidates are then history matched by the observation data. The application of these stochastic proxy models is essential, because it enables to search for the optimum solution in a wider range of the parameter space, therefore it avoids being trapped in local minima or losing the generality especially in the case of lacking field geological information. Unlike the conventional generation of prior geological models that require detailed petrophysical data, the stochastic proxy model uses the much coarser grid to represent the connectivity between the wells and reservoirs and incorporates the basic geological statistics of the field at the same time.

The application of the adjoint method and the coarse grid of the proxy model makes it possible to finish the history matching for a large ensemble of realizations within few hours. The trained proxy model can be further used in production optimization problems with high efficiency also because of its low degrees of freedom. The characterization of the geological information using a proxy model is another important extension compared with our previous work. This extension is demonstrated by testing on a three-phase 3D proxy model. It shows that the proxy approach can be used to characterize petrophysical information of the reservoir if the initial guess is close enough to a true model. Furthermore, we updated and modified the loss function by introducing the Dirac delta function. This makes the adjoint framework more robust because it solves the issue of the time measurement mismatch between the field observation and the simulator's model response, which occasionally happens because of the switching between the well controls and constraints.

2. Forward modeling formulation

DARTS is a high-performance numerical framework for modeling subsurface engineering problems. The physical kernels of DARTS utilize an operator-based linearization (OBL) technique (Voskov, 2017). This technique was proposed for addressing the challenges of balancing the accuracy of the numerical model and the performance of the simulator. The main idea of the OBL approach is parameterizing the state-dependent operators at the preprocessing stage that simplifies and accelerates the assembly of Jacobian. Later, the OBL technique was improved by adaptive parametrization (Khait and Voskov, 2018a) and implementation at GPU architecture (Khait et al., 2020). Currently, DARTS framework has been utilized for modeling of advanced petroleum (Khait and Voskov, 2018b; Lyu et al., 2021a), geothermal (Khait and Voskov, 2018c; Wang et al., 2020) and CO_2 sequestration (Kala and Voskov, 2020; Lyu et al., 2021b) applications.

2.1. General formulation of multiphase multicomponent flow

For forward modeling, the multiphase compositional flow problem with n_p phase and n_c components can be described by n_c equations of mass conservation:

$$\frac{\partial}{\partial t}\left(\phi \sum_{j=1}^{n_p} x_{cj}\rho_j s_j\right) + \mathrm{div}\sum_{j=1}^{n_p} x_{cj}\rho_j \boldsymbol{v}_j + \sum_{j=1}^{n_p} x_{cj}\rho_j \widetilde{q}_j = 0, \quad c = 1,\ldots,n_c, \quad (1)$$

where:

ϕ	porosity,
x_{cj}	mole fraction of component c in phase j,
s_j	phase saturation,
ρ_j	phase molar density,
$\boldsymbol{v}_j$	phase velocity,
$\widetilde{q}_j$	phase rate per unit volume.

For the phase velocity of each phase, Darcy's law is applied:

$$\boldsymbol{v}_j = -\left(\boldsymbol{K}\frac{k_{rj}}{\mu_j}\left(\nabla \boldsymbol{p}_j - \gamma_j \nabla \boldsymbol{d}\right)\right), j = 1,\ldots,n_p, \quad (2)$$

where:

$\boldsymbol{K}$	permeability tensor,
k_{rj}	relative permeability,
μ_j	phase velocity,
p_j	vector of pressures in phase j,
γ_j	gravity term,
$\boldsymbol{d}$	vector of depths (positive downwards).

Next, a finite-volume discretization on a general structured or unstructured mesh and backward Euler approximation in time are applied:

$$g = V\left(\left(\phi\sum_j x_{cj}\rho_j s_j\right)^{n+1} - \left(\phi\sum_j x_{cj}\rho_j s_j\right)^{n}\right) - \Delta t\sum_{l\in\boldsymbol{L}}\left(\sum_j x^l_{cj}\rho^l_j T^l_j \Delta\psi^l\right) + \Delta t\sum_j x_{cj}\rho_j q_j = 0, c = 1,\dots,n_c, \quad (3)$$

where V is control volume and $q_j = \tilde{q}_j V$ is the source of phase j. The effects of capillarity and gravity are neglected for simplicity. Two-Point Flux Approximation (TPFA) with upstream weighting introducing the summation over all interfaces $\boldsymbol{L}$ connecting the control volume with neighboring grid blocks are applied in the course of discretization. Based on these simplifications, $\Delta\psi^l$ becomes the pressure difference between two neighboring grid blocks, where T^l_j is phase transmissibility. These notations facilitate the further description of operator-based linearization.

Eq. (3) is the discretized form of flow and transport equations for general multi-component fluid. Another assumption of instantaneous thermodynamic equilibrium is chosen to close the system. Here, we used the overall molar formulation suggested by Collins et al. (1992). In the overall molar formulation, the unknowns are $[p, z_c]$ and the physical state $\boldsymbol{\omega}$ is completely defined by these variables. By applying the governing and closing relations, the values and their derivatives of all properties in Eq. (3) with respect to p and z_c can be obtained. To solve the resulting system of nonlinear equations, Newton–Raphson method is often applied in reservoir simulation. In this method, the linearized form of the nonlinear equation system is shown as:

$$\frac{\partial \mathbf{g}(\boldsymbol{\omega}^k)}{\partial \boldsymbol{\omega}^k}(\boldsymbol{\omega}^{k+1} - \boldsymbol{\omega}^k) = -\mathbf{g}(\boldsymbol{\omega}^k), \quad (4)$$

where k is the number of nonlinear iteration. The numerical approximations of the terms in Eq. (3) demand either reliable interpolation tables (e.g. standard PVT correlations), or a solution of the highly nonlinear equations (e.g. EoS-based properties).

2.2. Operator-Based Linearization

Based on OBL approach (Voskov, 2017), all terms in Eq. (3) can be defined as the function of two types of variables, which are physical state $\boldsymbol{\omega}$ and spatial coordinate $\boldsymbol{\xi}$. Then re-arrange Eq. (3) and define some new notations:

$$\boldsymbol{g}_c(\boldsymbol{\xi},\boldsymbol{\omega},\boldsymbol{w}) = V(\boldsymbol{\xi})\phi_0(\boldsymbol{\xi})\left(\alpha_c(\boldsymbol{\omega}) - \alpha_c(\boldsymbol{\omega^n})\right) - \Delta t\sum_l \beta^l_c(\boldsymbol{\omega})T^{ab}(\boldsymbol{\xi})(p^b - p^a) + \theta_c(\boldsymbol{\xi},\boldsymbol{\omega},\boldsymbol{w}) = 0,\ c = 1,\dots,n_c, \quad (5)$$

where:

$$\alpha_c(\boldsymbol{\omega}) = (1 + c_r(p - p_0))\sum_{j=1}^{n_p} x_{cj}\rho_j s_j, \quad (6)$$

$$\beta_c(\boldsymbol{\omega}) = \sum_{j=1}^{n_p} x_{cj}\frac{k_{rj}}{\mu_j}\rho_j, \quad (7)$$

$$\theta_c(\boldsymbol{\xi},\boldsymbol{\omega},\boldsymbol{w}) = \Delta t\sum_{j=1}^{n_p} x_{cj}\rho_j q_j(\boldsymbol{\xi},\boldsymbol{\omega},\boldsymbol{w}), \quad (8)$$

c_r is rock compressibility; ϕ_0 and p_0 are reference porosity and pressure; T^{ab} is transmissibility between block a and b; $\boldsymbol{\omega}$ and $\boldsymbol{\omega}^n$ are unknowns at the current and the previous timestep, respectively; and $\boldsymbol{w}$ are the variables defined well controls.

Note that α_c and β_c are only dependent on physical state $\boldsymbol{\omega}$. This indicates that both operators can be pre-processed and stored in tables for further use in the course of Jacobian and residual assembly. This pre-processing procedure largely reduces the computational cost and complexity in the simulation with highly nonlinear physics. To be more specific, instead of keeping track of each property and its derivatives with respect to nonlinear unknowns at each nonlinear iteration, the values of property can be efficiently reconstructed based on the pre-processed table defined in supporting points of physical state $\boldsymbol{\omega}$. When assembling the residual terms, a multilinear interpolation procedure is applied for the physical state $\boldsymbol{\omega}$ located inside of the hyper-cube defined by supporting points. The interpolation coefficients will yield partial derivatives needed for Jacobian construction which significantly simplify its assembly. Another advantage of this approach is that the abstract algebraic operators representing physics of any complexity shown in Eqs. (6)–(8) can be re-used many times in the process of forward simulation, data assimilation, or optimization.

3. Adjoint gradients formulation

In this section, we introduce the formulation of adjoint equations and gradients for history matching in OBL simulation framework. In our derivations, we will mostly follow the notations from Jansen (2011). The adjoint gradients formulation is implemented in DARTS framework for a wide range of energy transition applications.

In adjoint method, an augmented loss function $\bar{J}$ that shares the identical extrema with the original loss function J is constructed. It can be written as:

$$\bar{J}(\boldsymbol{\omega},\boldsymbol{u},\lambda) = \boldsymbol{J}(\boldsymbol{\omega},\boldsymbol{u}) + \lambda^{\mathrm{T}} g(\boldsymbol{\omega},\boldsymbol{u}), \quad (9)$$

where $\boldsymbol{\omega}$ is the state variables of the reservoir system, $\boldsymbol{u}$ is the control variables of the history matching problem, J is the loss function, g is the governing equation of the reservoir system, and λ^T is the transposed form of Lagrange multipliers. It is noticed that a new variable λ is introduced in the augmented loss function $\bar{J}(\boldsymbol{\omega},\boldsymbol{u},\lambda)$, compared with the original loss function $J(\boldsymbol{\omega},\boldsymbol{u})$. The extrema of Eq. (9), therefore, locate either at the boundary of the feasibility region or at stationary points. For the latter case, all first-order derivatives with respect to λ, $\boldsymbol{\omega}$ and $\boldsymbol{u}$ should be equal to zero, which leads to the following set of equations:

$$\bar{J}_\lambda = g(\boldsymbol{\omega},\boldsymbol{u}) = 0, \quad (10)$$

$$\bar{J}_{\boldsymbol{\omega}} = \lambda^{\mathrm{T}} g_{\boldsymbol{\omega}}(\boldsymbol{\omega},\boldsymbol{u}) + \boldsymbol{J}_{\boldsymbol{\omega}}(\boldsymbol{\omega},\boldsymbol{u}) = 0, \quad (11)$$

$$\bar{J}_{\boldsymbol{u}} = \lambda^{\mathrm{T}} g_{\boldsymbol{u}}(\boldsymbol{\omega},\boldsymbol{u}) + \boldsymbol{J}_{\boldsymbol{u}}(\boldsymbol{\omega},\boldsymbol{u}) = 0, \quad (12)$$

where the notations with subscript λ, $\boldsymbol{\omega}$ and $\boldsymbol{u}$ means the derivatives with respect to corresponding variables λ, $\boldsymbol{\omega}$ and $\boldsymbol{u}$, respectively. Eqs. (10)–(12) are the first-order necessary conditions for an optimum. The Eq. (10) is apparently satisfied already, because it is identical to the governing equations. Therefore, we only need to focus on Eqs. (11) and (12), which are known as adjoint equations and optimization equations, respectively.

For the reservoir simulation problem, we discretize its governing equation using finite volume method (FVM) spatially and backward Euler scheme temporally. The governing equation then becomes:

$$g_k(\boldsymbol{\omega}_k,\boldsymbol{\omega}_{k-1},\boldsymbol{u}) = \mathbf{0}, \quad (13)$$

where the subscript k and $(k-1)$ means the variables at the kth and $(k-1)$th time step. Generally, the loss function is dependent on the state variables $\boldsymbol{\omega}$ and control variables $\boldsymbol{u}$, which integrates nonlinear function j from time t_0 to t_K and can be written as:

$$\boldsymbol{J}(\boldsymbol{\omega},\boldsymbol{u}) = \int_{t_0}^{t_K} j(\boldsymbol{\omega},\boldsymbol{u},t)\mathrm{d}t. \quad (14)$$

Here the function j is usually nonlinearly dependent on the state variables $\boldsymbol{\omega}$ and control variables $\boldsymbol{u}$. It also represents the contribution to the loss function J at a given time t. This loss function in discretized form is written as:

$$J = \sum_{k=1}^{K} \boldsymbol{\Delta t}_k \boldsymbol{j}_k, \quad (15)$$

where Δt_k is the kth time step and K is the total number of the time steps of model response. The j_k is defined as:

$$j_k = \frac{1}{\Delta t_k}\delta_t(T_{obs}) \sum_{w=1}^{n_w}\sum_{j=1}^{n_p}\left(q_{w,j} - Q_{w,j}\right)^2, \tag{16}$$

where n_w is the total number of production wells; n_p is the total number of phases; $q_{w,j}$ and $Q_{w,j}$ are the model response and observation data of the wth production well and the jth phase at the time t, respectively; $\delta_t(T_{obs})$ is Dirac measure function and is given as:

$$\delta_t(T_{obs}) = \begin{cases} 1 & \text{if } t \in T_{obs} \\ 0 & \text{if } t \notin T_{obs} \end{cases}. \tag{17}$$

Note that δ_t is the function of T_{obs}, and t is actually the time at the endpoint of time interval Δt_k. T_{obs} is a set that only contains the marked time point in observation data, which means it is also a subset of all the time points of model response. The adjoint Eq. (11) in discretized form is given as:

$$\lambda_{k+1}^{\mathrm{T}}\frac{\partial g_{k+1}}{\partial \omega_k} + \lambda_k^{\mathrm{T}}\frac{\partial g_k}{\partial \omega_k} + \Delta t_k^{\mathrm{T}}\frac{\partial j_k}{\partial \omega_k} = 0, \tag{18}$$

$$\lambda_K^{\mathrm{T}}\frac{\partial g_K}{\partial \omega_K} + \Delta t_K^{\mathrm{T}}\frac{\partial j_K}{\partial \omega_K} = 0. \tag{19}$$

The Lagrange multipliers λ are found from the solution of Eqs. (18) and (19). Therefore, the gradient of loss function with respect to the control variables u can be written as:

$$\frac{\mathrm{d}J}{\mathrm{d}u} = \frac{\partial \bar{\mathscr{J}}}{\partial u} = \sum_{k=1}^{K}\left(\lambda_k^{\mathrm{T}}\frac{\partial g_k}{\partial u_k} + \Delta t_k^{\mathrm{T}}\frac{\partial j_k}{\partial u_k}\right). \tag{20}$$

Eq. (20) is the general form of adjoint gradients. For a specific reservoir engineering problem, the derivatives in this equation should be modified to corresponding expressions based on the selection of control variables. For example, when the control variables are not dependent on time, the adjoint gradient should be obtained by summing the term in Eq. (20) over all time steps, such as optimizing porosity or transmissibility (i.e. history matching). For the control variables are the function of time, we distribute the gradient terms of each time step into different control steps, and sum them up separately. The validation and comparison between the adjoint gradient and numerical gradient can be found in Tian et al. (2021).

4. Discrete Well Affinity model

In this section, we briefly describe important ingredients required for implementation of DiWA proxy methodology.

4.1. Connectivity graph

In reservoir numerical simulation, the governing equations of flow and transport in porous media need to be discretized temporally and spatially before solving the system of nonlinear equations. The spatial discretization of the reservoir is typically performed based on the control volume partitioning. This procedure provides a connectivity graph, which represents the spatial connections between discrete control volumes and associated transmissibilities (Lim, 1995). An automatic open-source meshing software GMSH (Geuzaine and Remacle, 2009) is utilized in this study to generate the connectivity graph of the proxy-model domain.

By setting different characteristic lengths in certain regions, GMSH is capable of meshing a geometric domain with different resolutions. When gridding the reservoir domain, we adjust the values of characteristic length to preserve the separation of wells in different control volumes while keeping the number of them low. The output file of GMSH contains information about vertexes, edges and volumes. Based on this information, an unstructured two-point discretization is applied to generate a connection list and associated parameters (e.g. transmissibility) based on the initial distribution of permeability and thickness (when known). The obtained model parameters can be used as an initial guess for the training. Here, the most important information required for the DiWA model is a connectivity graph which helps to evaluate wells affinity.

4.2. DiWA model construction

We utilize the Brugge field model to demonstrate the procedure of generating a DiWA model and to apply it in our framework. Brugge is a benchmark model for the optimization of reservoir production (Peters et al., 2010), which is used to generate observation data for the DiWA model training. The structure of the Brugge model consists of an East-West elongated half-dome with a large boundary fault at the northern edge. There are 20 production wells surrounded by 10 injection wells in this field. The realization encoded as FY-SS-KP-8-73 is used in this study to prepare the observation data of oil production. The permeability distribution of the realization FY-SS-KP-8-73 can be found in Fig. 1. This model runs for 3720 days with BHP control changing every 120 days. For the DiWA model, unstructured meshing and finite-volume discretization (Karimi-Fard et al., 2004) are applied to generate a coarse resolution DiWA model.

Notice that this unstructured DiWA model can only be considered as a proxy model since it is based on very basic information about the original field. For example, the reservoir boundaries are approximated by several piece-wise linear segments as can be seen in Fig. 1. The large boundary fault is not included in the unstructured DiWA model. The well locations are obtained based on projecting the real well locations to the nearest unstructured control volumes. The entire 2D domain of this field is meshed and then extruded in the vertical direction by an average constant thickness of the reservoir. Different characteristic lengths are set to generate a grid with coarser and finer meshing in the outer and inner boundaries, respectively. The finer inner resolution is explained by the main well locations and corresponding flow dynamics while the rest of the domain is located in the peripheral water drive with no significant flow. This helps to largely reduce the degrees of freedom of the proxy model. However, the DiWA model can always be further refined or elaborated simply by adding more piece-wise linear segments inside or at the boundary of the reservoir domain when it is necessary, for example, introducing the fault or extra cells around wells as shown in Tian et al. (2021).

As for the discretization of the well, it is conducted by adding the well to the connection list in the form of a well head block and a well body block. The well head, where the well control or constraint is assigned, is only connected with the well body, then the well body is further connected to the neighboring reservoir block. This procedure above shows the flexibility of discretizing the reservoir and wells, because we only need to update the existing connection list of the reservoir by adding two extra connectivities (i.e. the well head to the well body, and the well body to the reservoir block) for a well. One may also notice that the newly added well blocks may not be necessarily located in the center of the reservoir block, so that the well index cannot be calculated correctly. This problem can be solved by introducing an equivalent square block for the triangular cell where the well is located, see Eq. (32). Although this method cannot provide an accurate value of the well index for the given well, the approximated well index will be further updated in the course of training.

4.3. The pseudocode for the construction and application of diwa model

The construction procedure of DiWA model is shown in Algorithm 1. It should be noted that in this study, the sampling of the permeability K is based on the probability density of the high-fidelity model. However, the high-fidelity model is not the prerequisite in practice because this sampling procedure can be done as long as there is information of the probability density of the reservoir parameters (e.g., from logs).

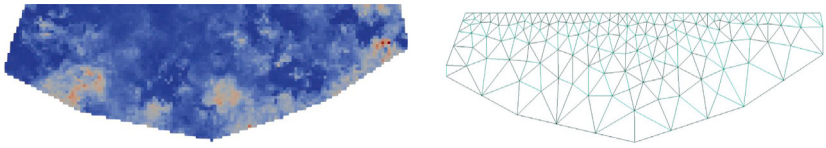


Fig. 1. The realization FY-SS-KP-8-73 and the unstructured DiWA model. Note that the realization FY-SS-KP-8-73 is a structured grid model with 44 550 active blocks, while there are only 283 cells in the unstructured DiWA model..

Once the collection of the DiWA model candidates are finished, they will be trained to fit the observation data by changing the values of the transmissibility T, well index T^w, and several parameters related to relative permeability. These parameters are the control variables in history matching and will be explained in Section 5. The trained DiWA models can be used to predict the reservoir production, because they are calibrated by the observation data. Production optimization is another application of these trained DiWA model. The pseudocode of the training and the forecast application of the DiWA models are shown in Algorithm 2.

Algorithm 1 The construction of DiWA model candidates

Use several piece-wise linear segments to represent the reservoir boundaries
Discretize the reservoir with very coarse unstructured 2D cells
Extrude the 2D mesh in vertical direction by an average constant reservoir thickness to form a 3D model
$seed = 1$, $N = 1$, and $list_{seed}$
while $N \leq 1000$ **do**
    Do the sampling of permeability K for each cell
    Generate the connection list to represent the connectivity between the neighboring cells
    Calculate the transmissibility T and add them to the connection list
    Add the well cells to the reservoir mesh
    Calculate the well index T^w for each well and add them to the connection list
    Run a forward simulation for the proxy model
    Calculate the misfit between the model response and the observation data
    if misfit $\leq$ threshold **then**
        Collect this model as candidate and save the $seed$ in $list_{seed}[N]$
        $N = N + 1$
    end if
    $seed = seed + 1$
end while

Algorithm 2 The training and the forecast application of the DiWA model

for N=1,2,3... **do**
    $seed \Leftarrow list_{seed}[N]$ ▷ $list_{seed}$ is generated from Algorithm 1
    Use $seed$ to reproduce the DiWA model candidate
    Re-scale T, T^w, and $\{S_{or},\ S_{wc},\ n_o,\ n_w,\ k^e_{rw}\rho_w/\mu_w,\ k^e_{ro}\rho_o/\mu_o\}$ in the range of (0, 1)
    while $misfit > tol_{misfit}$ and $stepsize > tol_{stepsize}$ **do**
        Calculate the gradients with respect to T and T^w using adjoint method
        Calculate the gradients with respect to $\{S_{or},\ S_{wc},\ n_o,\ n_w,\ k^e_{rw}\rho_w/\mu_w,\ k^e_{ro}\rho_o/\mu_o\}$ using finite difference method
        Update T, T^w, and $\{S_{or},\ S_{wc},\ n_o,\ n_w,\ k^e_{rw}\rho_w/\mu_w,\ k^e_{ro}\rho_o/\mu_o\}$ using the gradients
        Calculate the misfit between the model response and the observation data
    end while
    Set the total simulation time as $t = t_{training} + t_{forecast}$ to do the forecast
    Re-run the trained model using the optimized control variables
    Calculate the errors between the model response and the observation data
    Plot and compare the model response and the observation data
end for

5. Governing relations for control variables

Before training the model, it is essential to scale different control variables into suitable ranges of value. This is because some control variables may have an extremely different scale of values compared with other control variables in the optimization problem, and some of the existing optimization algorithms are sensitive to the scale of the optimization problem. There are some researches about the history matching using log transformation to the permeability when applying ensemble Kalman filter method (Evensen et al., 2007; Geir et al., 2002; Park and Choe, 2006). However, the log transformation is not applicable for the problem with highly non-Gaussian distribution (Shin et al., 2010; Zhou et al., 2011). In this study, history matching will be used in the Brugge field, which is not a reservoir with Gaussian distribution of permeability. Also, considering that there are three different types of control variables (transmissibility, well index, and rock-fluid interaction parameters) that will be used in the model training, we choose to apply the corresponding three different scaling factors to normalize them, instead of simply using the log transformation. For example, the magnitude of transmissibility is usually around 10000, while the connate water saturation and residual oil saturation are less than 0.49. The large difference between the magnitudes of different control variables may result in a very bad optimizer's performance. In this study, we choose to normalize them into the range of (0, 1). Therefore, different scaling factors need to be chosen for different types of control variables to make them fall into the same range of (0, 1). More details of the scaling factor can be found in Section 6. The gradients with respect to transmissibility and well index are calculated using the adjoint method to achieve high computational efficiency, while the gradients with respect to rock-fluid interaction parameters are calculated using numerical derivatives.

5.1. Transmissibility parameters

In the DARTS framework, connection list (Lim, 1995) is applied to represent the spatial connectivity graph of the reservoir model. The connection list contains information about the interfaces between grid blocks and the corresponding transmissibility of those connections. The transmissibility T^{ab} between grid blocks a and b is defined following (Karimi-Fard et al., 2004):

$$T^{ab} = c\left(\frac{\gamma_a\gamma_b}{\gamma_a+\gamma_b}\right),\ \gamma_i = A\frac{k_i}{D_i}, \tag{21}$$

where c is the unit conversion factor, A is the interface area between block a and b, D_i is the distance from the pressure node to the interface along the line connecting two pressure nodes and k_i is the grid block permeability.

It is noticed that T^{ab} can be considered as a linear parameter with respect to the flux term in Eq. (5), though it is nonlinear in respect to the loss function. The initial guess of transmissibility can be generated based on Eq. (21) with an appropriate estimate of petrophysical information. The number of transmissibility parameters is large in the proxy model, which corresponds to the high cost of gradient evaluation based on numerical gradients. In order to speed up these calculations, the adjoint method has been adopted for these parameters.

5.2. Well index parameters

Considering that the extremely coarse grid will be used to represent the complex real field, we introduce the well index as a control variable in the history matching. The well index can be seen as an additional degree of freedom to compensate for the extremely coarse scale of the control volumes. Well index is a measure of well potential to inject fluid into the reservoir or extract fluid from reservoir (Peaceman, 1983). For a single-phase flow, well index can be defined as:

$$T^w = c\frac{2\pi\Delta z\sqrt{k_x k_y}}{\ln r_o/r_w + S}, \tag{22}$$

where

$$r_o = 0.28\frac{\left[\left(k_y/k_x\right)^{1/2}\Delta x^2 + \left(k_x/k_y\right)^{1/2}\Delta y^2\right]^{1/2}}{\left(k_y/k_x\right)^{1/4} + \left(k_x/k_y\right)^{1/4}}. \tag{23}$$

Here Δx, Δy, and Δz are the block size in x, y, and z direction, respectively; k_x and k_y are the permeability in x and y direction, respectively; r_w is the radius of the well; S is skin factor.

It is noticed that T^w is also a linear parameter between the pressure gradient and the well rate. The amount of well index parameters is equal to the number of wells in the reservoir. In this study, the gradient evaluation with respect to well index parameters is also performed by the adjoint gradients approach.

5.3. Rock-fluid interaction parameters

For rock-fluid interactions, we use Brooks and Corey (1964) model to generate two-phase data. To cover the three-phase systems, we utilize the Stone 1 model (Stone, 1970) describing the relations of relative permeabilities between oil, water, and gas. In this model, the three-phase water relative permeability k_{rw} depends only on water saturation and is identical to k_{rwo} measured in the water/oil system:

$$k_{rw}(S_w) = k_{rwo}(S_w). \tag{24}$$

This is also true for relative permeability of gas in three-phase system:

$$k_{rg}(S_w) = k_{rgo}(S_g). \tag{25}$$

As for the relative permeability of oil k_{ro}, it depends on water and gas saturations:

$$k_{ro}(S_w, S_g) = \frac{1}{k_{row}(S_{wc})}S_o^*\frac{k_{row}(S_w)}{1-S_w^*}\frac{k_{rog}(S_g)}{1-S_g^*}, \tag{26}$$

where S_o^*, S_w^* and S_g^* are the scaled saturations of oil, water and gas, respectively:

$$S_o^* = \frac{S_o - S_{om}}{1 - S_{wc} - S_{om}}, \quad S_w^* = \frac{S_w - S_{wc}}{1 - S_{wc} - S_{om}}, \quad S_g^* = \frac{S_g}{1 - S_{wc} - S_{om}}. \tag{27}$$

According to Fayers and Matthews (1984), S_{om} is defined as:

$$S_{om} = \alpha S_{orw} + (1-\alpha)S_{org}, \quad \alpha = 1 - \frac{S_g}{1 - S_{wc} - S_{org}}. \tag{28}$$

where S_{orw} is the residual saturation of oil for a water/oil system, S_{org} is the residual saturation of oil for a gas/oil system.

The whole reservoir may be divided into several flow regions, where in a given region we assume that the fluid flow is governed by the same set of rock-fluid interaction parameters. For a water/oil system, the vector of these parameters are:

$$\upsilon_n = \left\{ n_w,\ n_o,\ S_{wc},\ S_{or},\ k^e_{rw}\rho_w/\mu_w,\ k^e_{ro}\rho_o/\mu_o \right\} \tag{29}$$

The first four parameters are dimensionless, and the last two parameters are scaling of β_c operator in Eq. (7). For a water/oil/gas system, this vector is written as:

$$\upsilon_n = \left\{ n_g,\ n_o,\ n_w,\ S_{gc}, S_{or},\ S_{wc},\ k^e_{rg},\ k^e_{ro},\ k^e_{rw} \right\} \tag{30}$$

The gradients with respect to the rock-fluid interaction parameters are calculated using numerical derivatives in the course of model training.

6. Results

In this section, we start with an example that proves the applicability of the DiWA approach for truly 3D models with a three-phase flow. Besides, this example demonstrates that the trained proxy model can recover the petrophysical information when the initial guess is close enough to the true solution. Finally, we will show an efficient approach for the generation of the DiWA proxy models which can be used in uncertainty quantification or robust optimization.

6.1. Deterministic reconstruction of permeabilities

To test the framework, we take a refined SPE 1 model (Odeh, 1981) for the generation of true data. In this model, there are mainly three different layers with constant permeability. Apart from the oil and water phase, the gas phase is also considered in this model and the gas is injected from the top layer at one of the reservoir corners. We build a coarse model and train it based on the observation data generated from the modified SPE 1 model. After the coarse model is trained, we interpret the regressed transmissibility back into permeability and compare them with the permeability distribution of the original SPE 1 model.

SPE 1 model is a basic test for the three-phase three-dimension black-oil reservoir modeling technique, see Fig. 2. In our modified version, there are 37500 grid blocks in total $[50 \times 50 \times 15]$. The block dimension in both x and y directions is 60 m. In z direction, each layer is divided into 5 grid blocks, and the block sizes for three layers in z direction are 1.2 m, 1.8 m, and 3 m, respectively. The porosity is 0.3. The permeabilities for three layers from top to bottom are 500 mD, 50 mD, and 200 mD, respectively. There are two wells (one injector and one producer) are located in the opposite corners of the reservoir. Gas is injected into the top layer and oil is produced from the bottom layer. The initial pressure of this model is 330 bar. The initial molar fraction of gas, oil, and water are 0.001, 0.648, and 0.351, respectively. The injector is set at a constant BHP control of 400 bar, and the producer is set at a constant BHP of 150 bar. The total simulation time of this model is 2000 days. More details about the parameters of PVT and relative permeability used in this model can be found in Appendix A.

A proxy model with 300 grid blocks $[10 \times 10 \times 3]$ is constructed and trained based on the observation data generated from the high-fidelity model. To make sure the well positions of the coarse model are identical to the high-fidelity model, we set the injector and producer at the grid block (1,1,1) and (10,10,3), respectively. Since three phases are considered in this problem, Stone 1 Model (Stone, 1970) is implemented to calculate the three-phase relative permeability, and the same sets of the parameters of the Stone 1 Model are applied in both high-fidelity

Table 1
The iterations and the errors between the rates of the DiWA and the high-fidelity model before and after training under different perturbations around the true transmissibility.

Perturbation	Error-initial, %	Error-trained, %	Iteration
2%	0.1167	0.0525	3
5%	0.0846	0.0501	4
10%	0.3057	0.0414	5
20%	2.0281	0.0436	2
40%	10.4036	0.0429	3
80%	45.8739	0.0450	12
100%	82.5188	0.0197	34

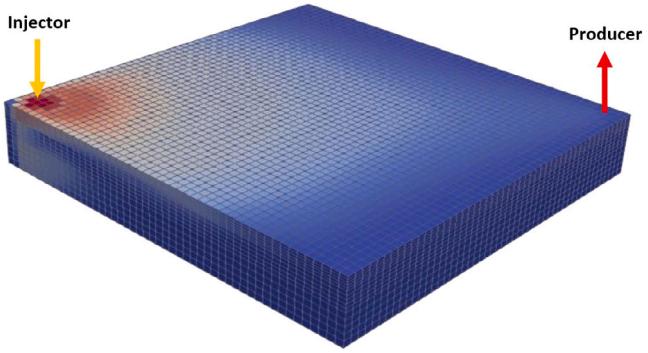


Fig. 2. SPE 1 model with two wells located at the opposite corners. The injector and its perforations are located at grid block (3,3,1–5). The producer and its perforations are located at grid block (48,48,11–15). The dimension in z direction is exaggerated 15 times for better visualization.

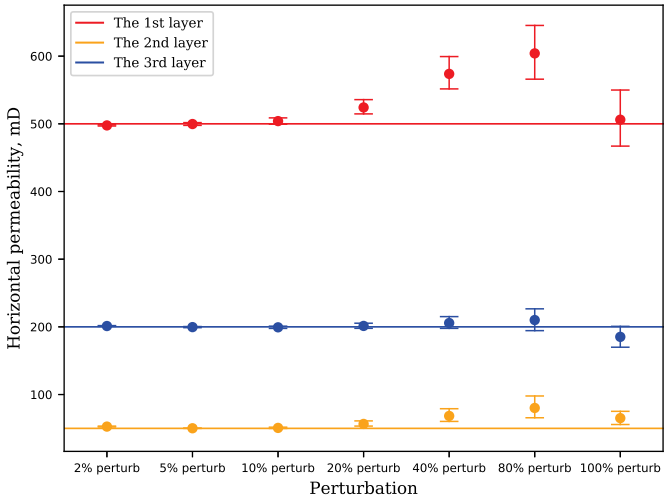


Fig. 3. The mean and the deviation of the horizontal permeability of DiWA model after training. The dots and the error bars represent the mean values and the deviation for the whole grid blocks in each layer, respectively.

and proxy model and they are not adjusted in the course of model training. Only the control variables of transmissibility and well index are adjusted in the course of model training using the adjoint gradients. The initial guess for transmissibilities and well indexes are generated by adding a random perturbation around the true transmissibility of the high-fidelity model. The scaling factors for transmissibilities and well indexes are 10000 and 1000, respectively. The error between the trained parameters and the true data is calculated using the following equation:

$$E = \frac{\delta_t(T_{obs}) \sum_{k=1}^{K} \sum_{w=1}^{n_w} \sum_{j=1}^{n_p} \left(q_{k,w,j} - Q_{k,w,j}\right)^2}{\sum_{k=1}^{K} \sum_{w=1}^{n_w} \sum_{j=1}^{n_p} \left(Q_{k,w,j}\right)^2}. \tag{31}$$

The training results are shown in Table 1 and Fig. 3.

It can be seen from Table 1 that as the increasing perturbation to the true transmissibility, the errors between the proxy and high-fidelity model increase from 0.1% to 83%. After the training, all proxy models can achieve a very small error value. It is noticed that the error after training for the 100%-perturbation case is 0.0197%, which is even smaller than the error of 0.0525% of the 2%-perturbation case. This can be explained by the non-uniqueness of the minimization procedure since the errors after training of all cases are close to zero.

However, a smaller error between the DiWA model and the high-fidelity model does not guarantee that the true solution is recovered due to the ill-posedness of the inverse problem. In Fig. 3, the mean values and deviations of the permeability values for each layer of the proxy model demonstrate significant divergence for the perturbation larger than 20%. This indicates that the selection of initial guesses can largely affect the characterization results of geological information. When the initial guess of the proxy model is generated based on highly reliable geological information, the trained model can recover the true petrophysical characteristics. However, when the geological information is missing or not sufficiently constrained, we need to apply a sampling procedure to generate the proxy model candidates. An example of the sampling procedure and training results is described next.

6.2. The training and production forecast of the diwa proxy model

We choose the unstructured model Case 1 and Case 5 from Tian et al. (2021) to test the proposed framework. Both grids are shown in Fig. 4. The control variables of the coarser proxy model consist of 394 transmissibility, 30 well indexes, and 6 rock-fluid interaction parameters. As for the finer proxy model, there are 953 transmissibility, 30 well indexes, and 6 rock-fluid interaction parameters.

We will first show the training and production forecast results based on the coarser model. The reservoir initial pressure, initial oil saturation, and BHP control in the DiWA model are kept identical to the high-fidelity model. Unlike the conventional process of the generation of prior geological models that requires detailed well logging data, seismic data, etc., we randomly sample the permeability for each cell of the DiWA model based on the probability density of the realization FY-SS-KP-8-73 shown in Fig. 5. The generation of these stochastic DiWA models can be done in the situation of very limited geological information, while it incorporates the basic geological statistics of the field.

Eq. (21) is then applied to compute the connection list and the corresponding transmissibilities for the coarser model. These transmissibilities are taken as the initial guess of transmissibility parameters and are bounded in the range of (0.01, 50 000). The scaling factor for the transmissibility is 50000. As for the initial guess for well indexes, Eqs. (22) and (23) provide an approximation of well index for the structured grid. Here we introduce an equivalent length of the triangular cell in the unstructured grid to replace the structured block size Δx and Δy in Eq. (23). They are given by:

$$\Delta x = \Delta y = \sqrt{A_{tri}}, \tag{32}$$

where A_{tri} is the area of the given triangular cell. Eq. (32) assumes that the triangular cell in the unstructured grid has the same effect on the

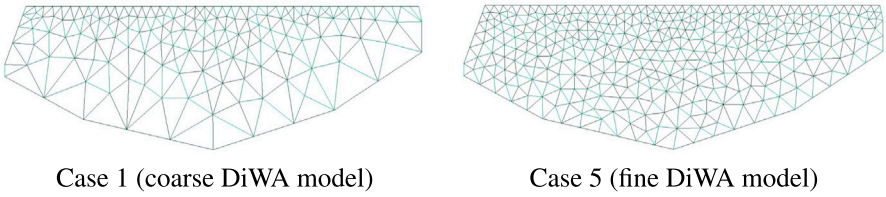


Fig. 4. Two DiWA models with different grid resolutions from Tian et al. (2021).

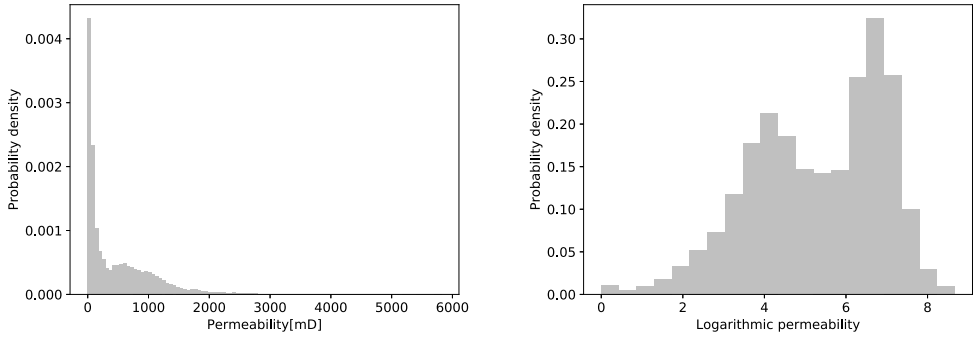


Fig. 5. The probability density for the permeability of the realization FY-SS-KP-8-73. The x axis of the right figure represents the logarithmic permeability.

well index as the square block with the same area in the structured grid. The well index is bounded in the range of (0.001, 10 000), and they are normalized by the scaling factor 10000. The initial guess of 6 rock-fluid interaction parameters are uniformly sampled from the range of (0.00001, 5), (0.00001, 5), (0, 0.49), (0, 0.49), (100, 3000), and (10, 2000), respectively. These 6 rock-fluid interaction parameters will be applied in the whole reservoir, and they are normalized by the scaling factors 5, 5, 0.49, 0.49, 3000, and 2000, respectively. The parameters of relative permeability and the PVT data of oil used in the Brugge high-fidelity model can be found in Appendix B.

The high-fidelity model runs for 3720 days to generate true observation data. For every 120 days, the BHP controls of injection wells are set as 170 bar plus uniformly distributed random perturbations ranging from 1 to 30, and the BHP controls of production wells are set 130 bar minus uniformly distributed random perturbations ranging from 1 to 30, see Fig. C.14 in Appendix C. We choose the first 3000 days as the training period to train the DiWA model. After that, the DiWA model continues to run for 720 more days in order to test the accuracy of the forecasting period compared with the true response.

Before the training of the DiWA model, we first sample for many different realizations and then check their misfit of the total oil and water rate between the model response and observation data. The threshold for the rate variability is taken as 60% of true rate, and we use this threshold to filter out the outliers. Only the realization that meets this threshold will be retained otherwise discarded. This sampling procedure would not stop until 1000 candidates of prior realization are collected. Later, these 1000 realizations are set as the initial guesses to train the DiWA model. Note that the training procedure for the ensemble of proxy models can be finished within a limited time thanks to the high efficiency of gradient calculation using the adjoint method and the utilization of computer clusters. The total training time of the 1000 realizations took below 12 hours on 10 cluster nodes with total 40 Intel Xeon CPU E5-2650 v3 processors.

The results of the total oil and water production rate of 100 best realizations are shown in Fig. 6. The gray and blue curves represent the rate before and after training, respectively. The red curves show the true data. The vertical dashed lines separate the training periods and the prediction periods. It is clearly seen that proxy models cannot capture the characteristics of the first time period (i.e. transient period that is completed at 400 days) of oil production due to the coarse representation of well connectivity in the proxy methodology. However, the later period is captured quite well and the deviation of the model response from the true solution is very limited. This deviation increases when the model switch from training to forecast period which is expected.

More detailed comparisons of oil rates for 20 production wells are plotted in Fig. 7. The results show that the mean error of the training period and prediction period of oil rate are 0.59% and 0.62%, respectively. As it can be seen from Fig. 7, the oil rates of production wells have a good match with the true data in the training period (before the vertical dashed lines). It can be seen that only a few production wells have a relatively larger spread of oil rates in the forecast period compared with the spread in the training period. The water production curves of these 20 production wells are shown in Fig. 8. As it can be seen that most of the wells also have a good match with the true data of water production rates. There are three wells (Well:P01, P04, and P05) that have water breakthrough after the training period. We plotted these three wells in Fig. 9 separately. Note that the magnitudes of y axis (water production rate) of these three wells are smaller than the rest of the wells. Most of the realizations capture the timing of the water breakthrough. But the water rates of these realizations after water breakthrough have a relatively larger spread compared with the curves before water breakthrough. Again, these results are expected since proxy models have a limited forecast capability.

Similarly, we implement the identical filtering and training strategy for the finer model in Fig. 4, which is a much finer unstructured grid and has more degrees of freedom compared with the coarser model. The total production rates of oil and water are shown in Fig. 10. The oil rates of 20 production wells are plotted in Fig. 11. It can be seen that the spread of the oil rates has been reduced for some wells. The resulting mean error of the training period and prediction period are 0.44% and 0.17%, respectively. The total training time took below 23 h on the same cluster setup. The water production curves of these 20 production wells are plotted in Fig. 12. The forecast results of water breakthrough of Well:P01, P04, and P05 are shown in Fig. 13. Similar conclusion can be drawn that the spread of the water rates is reduced compared with the results of Case 1 in Fig. 8, and most of the realizations are able to capture the timing of water breakthrough. The forecast after water breakthrough also has a relatively larger spread compared with the curves before water breakthrough.

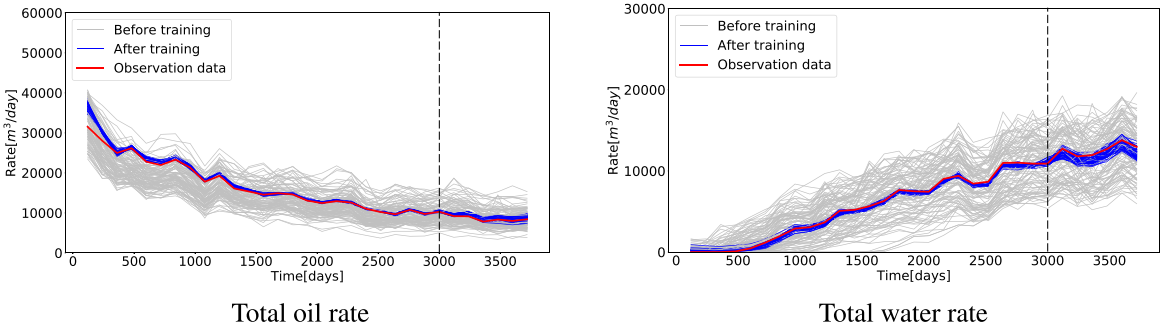


Fig. 6. The total oil (left) and water (right) production rate of 100 best fit realizations before and after training of Case 1 (coarse model).

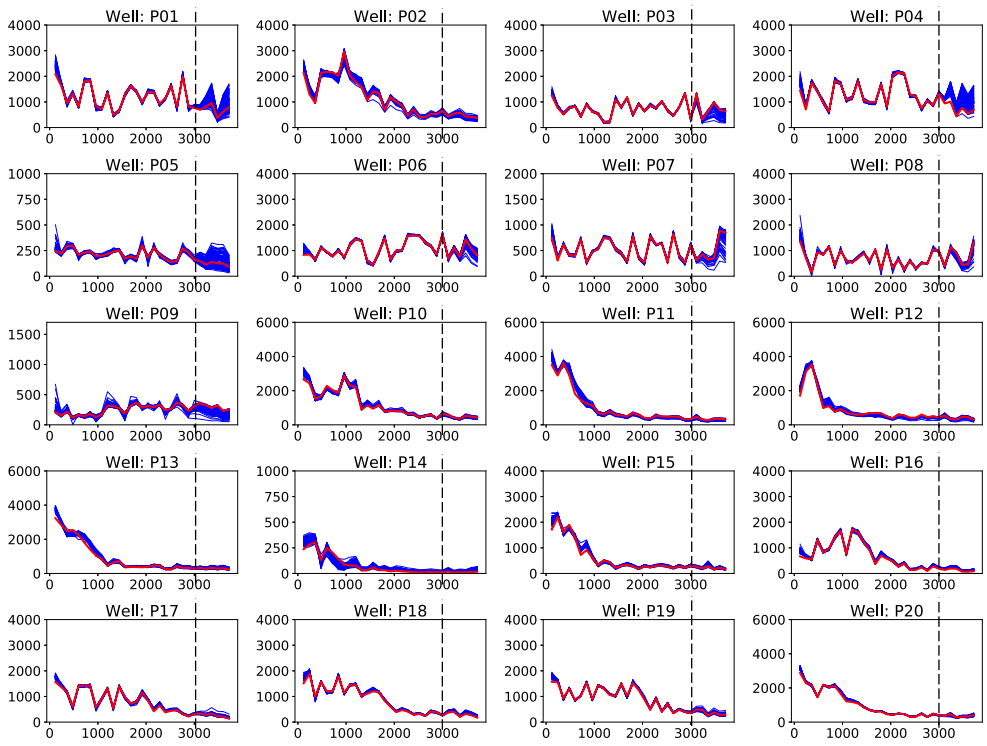


Fig. 7. The oil rates of the 100 best fit realizations (blue curves) and the observation data (red curves); x axis shows the production time (days), y axis shows the oil production rate (m^3/day); the mean error of training period and prediction period between these 100 realizations of DiWA model of Case 1 (coarse model) and the observation data are 0.59% and 0.62%, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

7. Conclusion and discussion

In this study, we presented a new adjoint framework using an Operator-Based Linearization (OBL) approach. This framework takes the advantage of the high efficiency of the OBL technique and exploits an abstract interpretation of physics benefiting computational performance for a forward simulation. The adjoint framework is implemented in the Delft Advanced Research Terra Simulator (DARTS) and applied for calculating the gradients with respect to transmissibility and well indexes. Later, these adjoint gradients and the numerical derivatives with respect to the rock-fluid parameters are intergrated for the training of proxy models based on gradient optimization. We constructed Discrete Well Affinity (DiWA) data-driven proxy model with a limited number of degrees of freedom and impressive forward simulation performance. The main idea of the DiWA model is that it uses very coarse grid and very basic geological information to represent a complex reservoir structure and the fluid flow in the reservoir. The coarse grid can be either structured or unstructured. DiWA model includes the geological properties like the average reservoir thickness, porosity, etc. But more geological properties can also be added to the DiWA model if they are available.

In this work, we tested a simple 3D proxy model with multiple layers and three flowing phases. This model was trained to match observation data generated from a modified SPE 1 model. The results showed that the proxy model can be used to characterize geological information of the reservoir when the initial guess is generated based on reliable geological information. But this may not be true when the initial guess for the model is far from the true geology.

This inspires us to introduce an efficient sampling approach where each statistical member has been trained to match the true data. A high-fidelity Brugge model was utilized to generate true data for model training and test the performance of the proposed approach. In the procedure of true data generation, uniformly distributed random perturbations were added to BHP control. The generated true data were used then to train the DiWA model. The proxy model used in this study only contains some basic geological information like the contour and the average thickness of the original reservoir model. However, more information can be added to this framework to improve the training process with appropriate regularization.

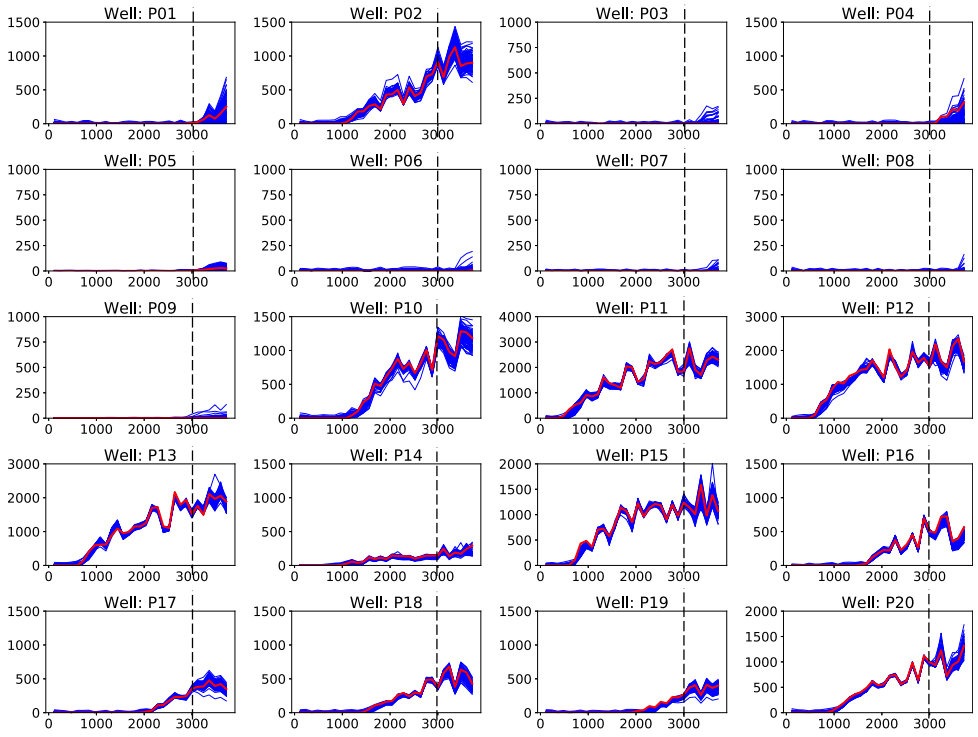


Fig. 8. The water production rates of the 100 best fit realizations (blue curves) and the observation data (red curves) for Case 1; x axis shows the production time (days), y axis shows the oil production rate (m^3/day). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

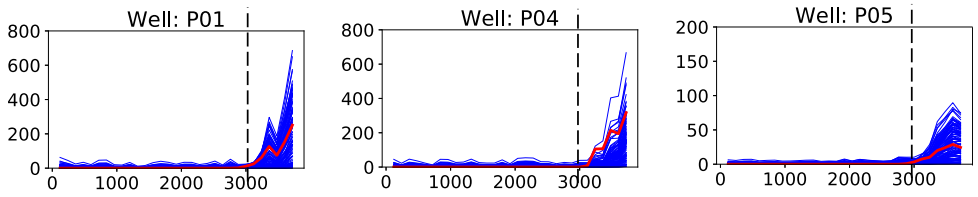


Fig. 9. The water production rates of the 100 best fit realizations (blue curves) and the observation data (red curves) for Well:P01, P04 and P05 of Case 1; x axis shows the production time (days), y axis shows the oil production rate (m^3/day). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

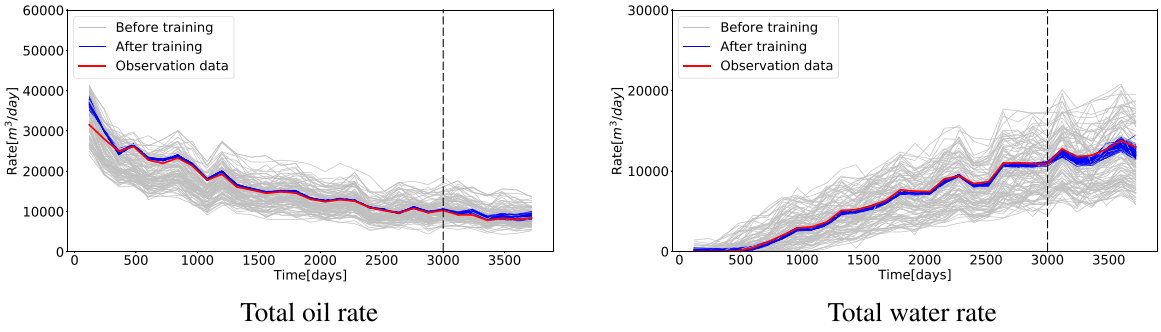


Fig. 10. The total oil (left) and water (right) production rate of 100 best fit realizations before and after training of Case 5 (fine model).

The proposed framework has several benefits. Firstly, the OBL technique helps to increase the efficiency of the assembly of the residual and the Jacobian in both forward simulation and adjoint gradients calculation. Secondly, the performance of the training improved significantly since the cost of the adjoint gradient calculation is almost equivalent to a single forward run, which makes feasible a proposed stochastic DiWA proxy methodology. Furthermore, if more control variables are introduced to the model, a higher convergence efficiency of the adjoint method can be obtained compared with the numerical gradients approach.

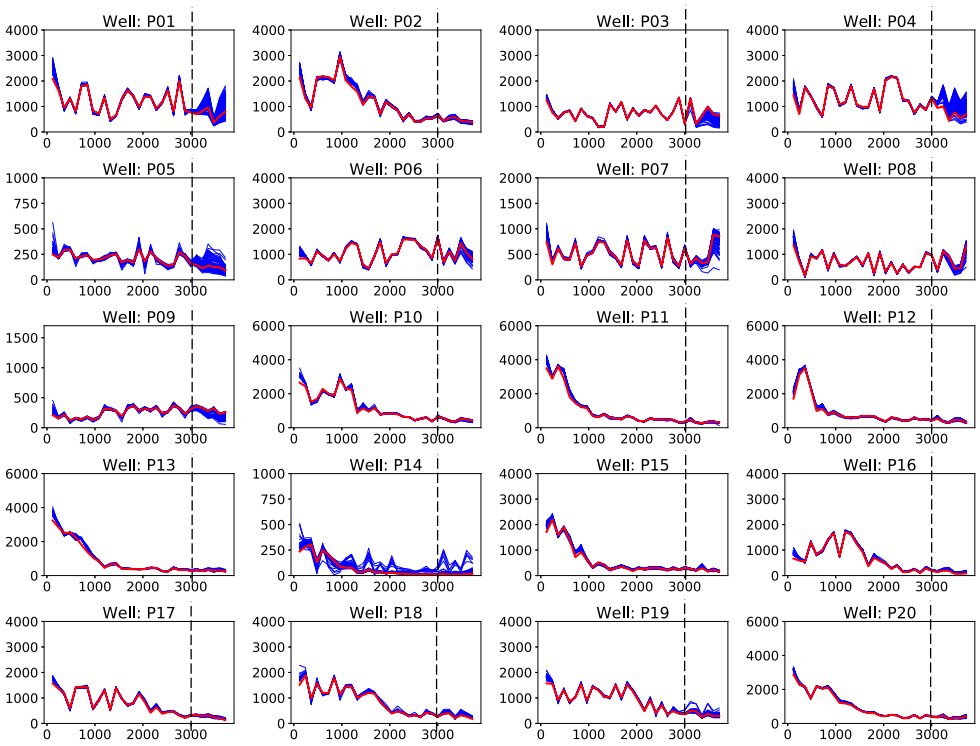


Fig. 11. The oil rates of the 100 best fit realizations (blue curves) and the observation data (red curves); x axis shows the production time (days), y axis shows the oil production rate (m^3/day); the mean error of training period and prediction period between these 100 realizations of DiWA model of Case 5 (fine model) and the observation data are 0.44% and 0.17%, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

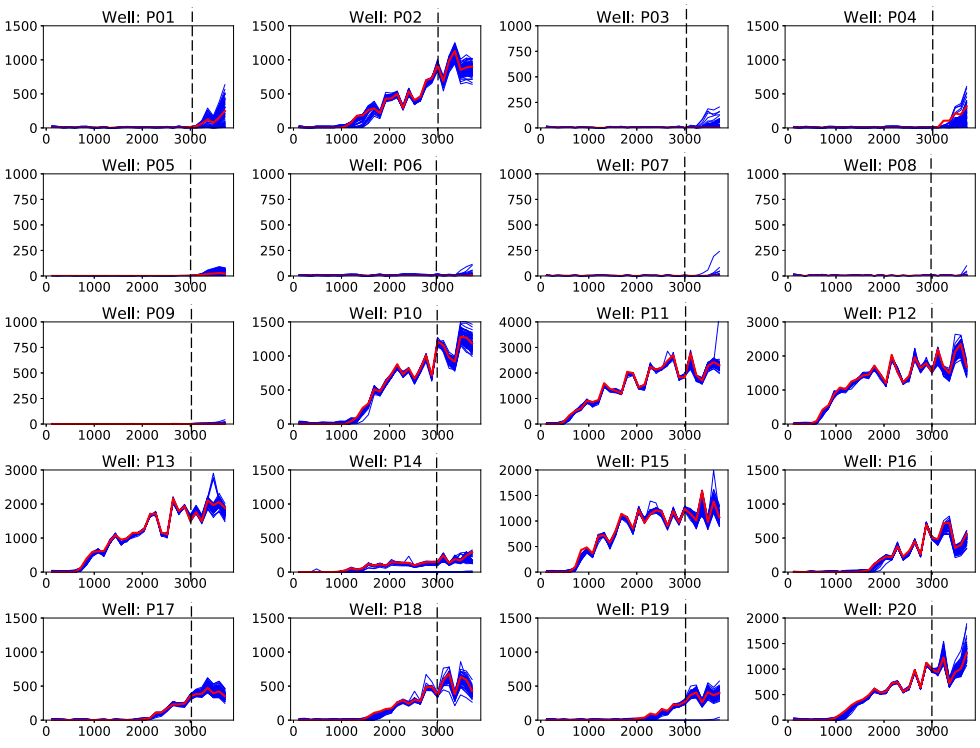


Fig. 12. The water production rates of the 100 best fit realizations (blue curves) and the observation data (red curves) for Case 5; x axis shows the production time (days), y axis shows the oil production rate (m^3/day). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

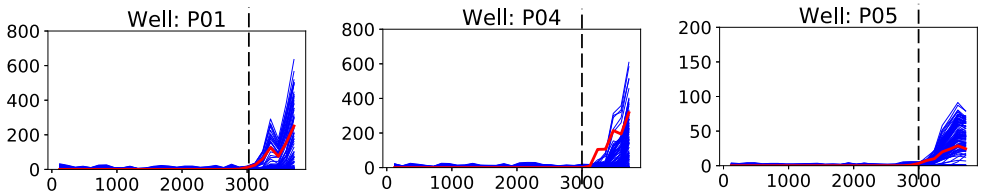


Fig. 13. The water production rates of the 100 best fit realizations (blue curves) and the observation data (red curves) for Well:P01, P04 and P05 of Case 5; x axis shows the production time (days), y axis shows the oil production rate (m^3/day). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

CRediT authorship contribution statement

Xiaoming Tian: Methodology, Software, Writing – original draft, Writing – review & editing. **Denis Voskov:** Conceptualization, Methodology, Software, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to thank Oleg Volkov and Jan Dirk Jansen for their valuable suggestions during the course of this project.

Appendix A

SPE 1 model is a three-phase three-dimensional black-oil model from the SPE-9723-PA paper (Odeh, 1981), which means there are dissolved gas existing in the reservoir fluids. The density of oil, water, and gas at the surface condition are 786.5 kg/m^3, 1037.8 kg/m^3, and 0.97 kg/m^3, respectively. The rock compressibility is 4.35×10^{-5} bar^{-1} when the reference pressure is 277.0 bar. The water formation volume factor, compressibility, and viscosity at the same reference pressure are

Table A.2
The parameters of relative permeability used in SPE 1 model.

Phase	Oil	Water	Gas
Residual saturation	0.103	0.197	0.013
End point relative permeability	0.584	0.910	0.830
Saturation exponent	2.54	1.01	1.62

Table A.3
The formation volume factor and viscosity of the gas phase in SPE 1 model.

Pressure (bar)	Formation volume factor (–)	Viscosity (cp)
1	166.666	0.008
18	12.093	0.0096
35	6.274	0.0112
69	3.197	0.014
137	1.614	0.0189
171	1.294	0.0208
205	1.08	0.0228
273	0.811	0.0268
341	0.649	0.0309
613	0.386	0.047

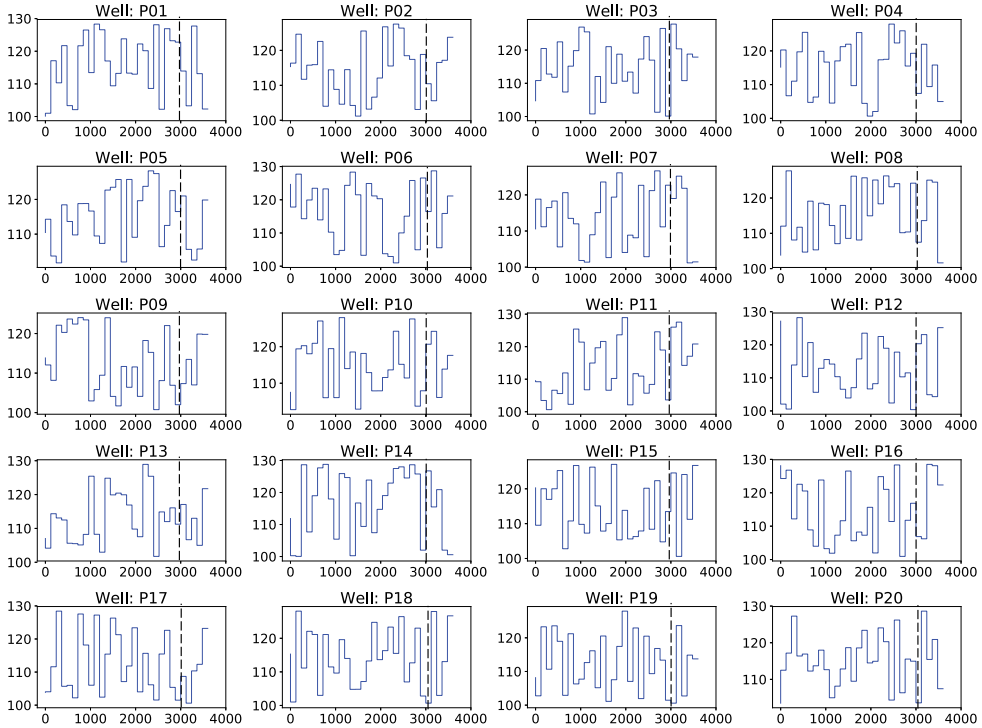


Fig. C.14. The BHP control of production wells; x axis shows the production time (days), y aixs shows the BHP of production wells (bar); The vertical dashed lines separate the training periods and the prediction periods.

Table A.4
The formation volume factor, gas oil ratio, and viscosity of the oil phase in SPE 1 model.

Pressure (bar)	Formation volume factor (–)	Gas oil ratio (–)	Viscosity (cp)
1	1.062	0.001	1.04
18	1.15	0.0905	0.975
35	1.207	0.18	0.91
69	1.295	0.371	0.83
137	1.435	0.636	0.695
171	1.5	0.775	0.641
205	1.565	0.93	0.594
273	1.695	1.270	0.51
613	1.579	1.270	0.74

Table B.5
The parameters of relative permeability used in Brugge high-fidelity model.

Phase	Oil	Water
Residual saturation	0.15	0.225
End point relative permeability	0.4	1.0
Saturation exponent	3.0	3.0

Table B.6
The formation volume factor and viscosity of oil in the Brugge model.

Pressure (bar)	Formation volume factor (–)	Viscosity (cp)
1	1.0007	1.294
170	0.9780	1.294
500	0.9337	1.294

1.029, 4.54×10^{-5} bar^{-1}, and 0.31 cp, respectively. The parameters related to the relative permeability and PVT data of oil and gas phases can be found in Tables A.2–A.4.

Appendix B

The dead oil model is applied in the Brugge field. There are only oil and water phases existing in the reservoir throughout the production period. The density of oil and water are 897.0 kg/m^3 and 1002.8 kg/m^3, respectively. The rock compressibility is 5.08×10^{-5} bar^{-1} when the reference pressure is 170.0 bar. The water formation volume factor, compressibility, and viscosity at the same reference pressure are 0.9927, 4.35×10^{-5} bar^{-1}, and 0.32 cp, respectively. The parameters related to the relative permeability and PVT data of oil can be found in Tables B.5 and B.6

Appendix C

See Fig. C.14

References

Batycky, R., Blunt, M., Thiele, M., 1997. A 3D field-scale streamline-based reservoir simulator. SPE Reserv. Eng. 12 (04), 246–254. http://dx.doi.org/10.2118/36726-pa.

Brooks, R., Corey, A., 1964. Hydraulic properties of porous media. Hydrol. Pap. 3.

Brouwer, D., Jansen, J., 2004. Dynamic optimization of waterflooding with smart wells using optimal control theory. SPE J. 9 (04), 391–402. http://dx.doi.org/10.2118/78278-pa.

Bryson, A.E., Ho, Y.-C., 2018. Applied Optimal Control. Routledge, http://dx.doi.org/10.1201/9781315137667.

Collins, D., Nghiem, L., Li, Y.-K., Grabonstotter, J., 1992. An efficient approach to adaptive- implicit compositional simulation with an equation of state. SPE Reserv. Eng. 7 (02), 259–264. http://dx.doi.org/10.2118/15133-pa.

Durlofsky, L., 2005. Upscaling and Gridding of Fine Scale Geological Models for Flow Simulation, in: The 8th International Forum on Reservoir Simulation Iles Borromees, Stresa, Italy.

Eberhart, R., Kennedy, J., 1995. A new optimizer using particle swarm theory. In: MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science. IEEE, http://dx.doi.org/10.1109/mhs.1995.494215.

Evensen, G., Hove, J., Meisingset, H., Reiso, E., Seim, K.S., Espelid, Ø., 2007. Using the EnKF for Assisted History Matching of a North Sea Reservoir Model. SPE, http://dx.doi.org/10.2118/106184-ms.

Fathi, Z., Ramirez, F.W., 1984. Optimal injection policies for enhanced oil recovery: Part 2-surfactant flooding. Soc. Petrol. Eng. J. 24 (03), 333–341. http://dx.doi.org/10.2118/12814-pa.

Fayers, F., Matthews, J., 1984. Evaluation of normalized stone's methods for estimating three-phase relative permeabilities. Soc. Petrol. Eng. J. 24 (02), 224–232. http://dx.doi.org/10.2118/11277-pa.

Geir, Æ., Mannseth, T., Vefring, E.H., 2002. Near-Well Reservoir Monitoring Through Ensemble Kalman Filter. SPE, http://dx.doi.org/10.2118/75235-ms.

Geuzaine, C., Remacle, J.-F., 2009. Gmsh: A 3-D finite element mesh generator with built-in pre- and post-processing facilities. Internat. J. Numer. Methods Engrg. 79 (11), 1309–1331. http://dx.doi.org/10.1002/nme.2579.

Guo, Z., Reynolds, A.C., Zhao, H., 2018. A physics-based data-driven model for history matching, prediction, and characterization of waterflooding performance. SPE J. 23 (02), 367–395. http://dx.doi.org/10.2118/182660-pa.

Holland, J.H., 1984. Genetic algorithms and adaptation. In: Adaptive Control of Ill-Defined Systems. Springer US, pp. 317–333. http://dx.doi.org/10.1007/978-1-4684-8941-5_21.

Jansen, J., 2011. Adjoint-based optimization of multi-phase flow through porous media - a review. Comput. & Fluids 46 (1), 40–51. http://dx.doi.org/10.1016/j.compfluid.2010.09.039.

Jansen, F., Kelkar, M., 1997. Non-stationary estimation of reservoir properties using production data. In: SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers, http://dx.doi.org/10.2118/38729-ms.

Jenny, P., Lee, S., Tchelepi, H., 2003. Multi-scale finite-volume method for elliptic problems in subsurface flow simulation. J. Comput. Phys. 187 (1), 47–67. http://dx.doi.org/10.1016/s0021-9991(03)00075-5.

Kala, K., Voskov, D., 2020. Element balance formulation in reactive compositional flow and transport with parameterization technique. Comput. Geosci. 24 (2), 609–624. http://dx.doi.org/10.1007/s10596-019-9828-y.

Karimi-Fard, M., Durlofsky, L., Aziz, K., 2004. An efficient discrete-fracture model applicable for general-purpose reservoir simulators. SPE J. 9 (02), 227–236. http://dx.doi.org/10.2118/88812-pa.

Khait, M., Voskov, D., 2018a. Operator-based linearization for efficient modeling of geothermal processes. Geothermics 74, 7–18. http://dx.doi.org/10.1016/j.geothermics.2018.01.012.

Khait, M., Voskov, D., 2018b. Adaptive parameterization for solving of thermal/compositional nonlinear flow and transport with buoyancy. SPE J. 23 (2), 522–534. http://dx.doi.org/10.2118/182685-pa.

Khait, M., Voskov, D., 2018c. Operator-based linearization for efficient modeling of geothermal processes. Geothermics 74, 7–18. http://dx.doi.org/10.1016/j.geothermics.2018.01.012.

Khait, M., Voskov, D., Zaydullin, R., 2020. High performance framework for modelling of complex subsurface flow and transport applications. In: ECMOR 2020 - 17th European Conference on the Mathematics of Oil Recovery. http://dx.doi.org/10.3997/2214-4609.202035188.

Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P., 1983. Optimization by simulated annealing. Science 220 (4598), 671–680. http://dx.doi.org/10.1126/science.220.4598.671.

Kourounis, D., Voskov, D., Aziz, K., 2010. Adjoint methods for multicomponent flow simulation. In: ECMOR 2010 - 12th European Conference on the Mathematics of Oil Recovery. http://dx.doi.org/10.3997/2214-4609.20144996.

Lim, K.-T., 1995. A new approach for residual and Jacobian arrays construction in reservoir simulators. SPE Comput. Appl. 7 (04), 93–96. http://dx.doi.org/10.2118/28248-pa.

Lyu, X., Khait, M., Voskov, D., 2021a. Operator-based linearization approach for modelling of multiphase flow with buoyancy and capillarity. SPE J. http://dx.doi.org/10.2118/205378-PA.

Lyu, X., Voskov, D., Rossen, W.R., 2021b. Numerical investigations of foam-assisted CO_2 storage in saline aquifers. Int. J. Greenhouse Gas Control http://dx.doi.org/10.1016/j.ijggc.2021.103314.

Mehos, G.J., Ramirez, W., 1989. Use of optimal control theory to optimize carbon dioxide miscible-flooding enhanced oil recovery. J. Pet. Sci. Eng. 2 (4), 247–260. http://dx.doi.org/10.1016/0920-4105(89)90002-8.

Moraes, R., Rodrigues, J., Hajibeygi, H., Jansen, J., 2017. Multiscale gradient computation for flow in heterogeneous porous media. J. Comput. Phys. 336, 644–663. http://dx.doi.org/10.1016/j.jcp.2017.02.024.

Odeh, A.S., 1981. Comparison of solutions to a three-dimensional black-oil reservoir simulation problem (includes associated paper 9741). J. Pet. Technol. 33 (01), 13–25. http://dx.doi.org/10.2118/9723-pa.

Park, K., Choe, J., 2006. Use of Ensemble Kalman Filter to 3-Dimensional Reservoir Characterization during Waterflooding. SPE, http://dx.doi.org/10.2118/100178-ms.

Peaceman, D., 1983. Interpretation of well-block pressures in numerical reservoir simulation with nonsquare grid blocks and anisotropic permeability. Soc. Petrol. Eng. J. 23 (03), 531–543. http://dx.doi.org/10.2118/10528-pa.

Peters, L., Arts, R., Brouwer, G., Geel, C., Cullick, S., Lorentzen, R.J., Chen, Y., Dunlop, N., Vossepoel, F.C., Xu, R., Sarma, P., Alhuthali, A.H., Reynolds, A., 2010. Results of the brugge benchmark study for flooding optimization and history matching. SPE Reserv. Eval. Eng. 13 (03), 391–405. http://dx.doi.org/10.2118/119094-pa.

Ramirez, W.F., Fathi, Z., Cagnol, J.L., 1984. Optimal injection policies for enhanced oil recovery: Part 1 theory and computational strategies. Soc. Petrol. Eng. J. 24 (03), 328–332. http://dx.doi.org/10.2118/11285-pa.

Sarma, P., Aziz, K., Durlofsky, L., 2005. Implementation of adjoint solution for optimal control of smart wells. In: SPE Reservoir Simulation Symposium. Society of Petroleum Engineers, http://dx.doi.org/10.2118/92864-ms.

Sarma, P., Durlofsky, L.J., Aziz, K., Chen, W.H., 2006. Efficient real-time reservoir management using adjoint-based optimal control and model updating. Comput. Geosci. 10 (1), 3–36. http://dx.doi.org/10.1007/s10596-005-9009-z.

Shin, Y., Jeong, H., Choe, J., 2010. Reservoir Characterization Using an EnKF and a Non-parametric Approach for Highly Non-Gaussian Permeability Fields, 32 (16) 1569–1578, http://dx.doi.org/10.1080/15567030902804780.

Stone, H., 1970. Probability model for estimating three-phase relative permeability. J. Pet. Technol. 22 (02), 214–218. http://dx.doi.org/10.2118/2116-pa.

Tian, X., Blinovs, A., Khait, M., Voskov, D., 2021. Discrete well affinity (DiWA) data-driven proxy model for production forecast. SPE J. 1–17. http://dx.doi.org/10.2118/205489-pa.

Volkov, O., Voskov, D., 2016. Effect of time stepping strategy on adjoint-based production optimization. Comput. Geosci. 20 (3), 707–722. http://dx.doi.org/10.1007/s10596-015-9528-1.

Voskov, D.V., 2017. Operator-based linearization approach for modeling of multiphase multi-component flow in porous media. J. Comput. Phys. 337, 275–288. http://dx.doi.org/10.1016/j.jcp.2017.02.041.

Wang, Y., Voskov, D., Khait, M., Bruhn, D., 2020. An efficient numerical simulator for geothermal simulation: A benchmark study. Appl. Energy 264, http://dx.doi.org/10.1016/j.apenergy.2020.114693.

Wei, L., Ramirez, W.F., Qi, Y.F., 1993. Optimal control of steamflooding. SPE Adv. Technol. Ser. 1 (02), 73–82. http://dx.doi.org/10.2118/21619-pa.

Zhou, H., Li, L., Franssen, H.-J.H., Gómez-Hernández, J.J., 2011. Pattern Recognition in a Bimodal Aquifer Using the Normal-Score Ensemble Kalman Filter, 44 (2) 169–185, http://dx.doi.org/10.1007/s11004-011-9372-3.

Zubarev, D.I., 2009. Pros and cons of applying proxy-models as a substitute for full reservoir simulations. In: SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers, http://dx.doi.org/10.2118/124815-ms.