

Original article

A general physics-based data-driven framework for numerical simulation and history matching of reservoirs

Xiang Rao^{1,2}*, Yunfeng Xu^{1,2}*, Deng Liu^{1,2}, Yina Liu^{1,2}, Yujie Hu^{1,2}

¹*School of Petroleum Engineering, Yangtze University, Wuhan 430100, P. R. China*

²*Cooperative Innovation Center of Unconventional Oil and Gas (Ministry of Education & Hubei Province), Yangtze University, Wuhan 430100, P. R. China*

Keywords:

Data-driven model
physics-based model
numerical simulation
history matching

Cited as:

Rao, X., Xu, Y., Liu, D., Liu, Y., Hu, Y.
A general physics-based data-driven
framework for numerical simulation and
history matching of reservoirs. *Advances
in Geo-Energy Research*, 2021, 5(4):
422-436, doi: 10.46690/ager.2021.04.07

Abstract:

This paper proposed a general physics-based data-driven framework for numerical modeling and history matching of reservoirs that achieves a good balance of flow physics and actual field data. Underground reservoir is easily discretized in this framework as a flow network composed of one-dimensional connection elements, each of which is defined by two flow characteristic parameters. Each one-dimensional connection element is divided into some grids, and the cross-sectional area and permeability of the grids on the same connection element are equal. The fully implicit scheme of flow equations and the Newton iteration nonlinear solver concurrently solve all unknown quantities. Then, using actual field data, the simultaneous perturbation stochastic approximation algorithm is used to invert flow characteristic parameters of each connection element, and the unequal constraint that the volume of connection elements should not exceed the total reservoir volume is added to control the data-driven process. To demonstrate the unequal constraint is physical, a test case of a waterflooding reservoir with a high permeability zone is given. A waterflooding reservoir example with five injectors and four producers is used to demonstrate that this framework outperforms earlier techniques, and another case with single-phase depletion development is used to demonstrate that this framework has a high generalization for flow models. In addition, this data-driven framework based on physics is expected to serve as a reference for other fields of science and engineering.

1. Introduction

Only a few static and dynamic data collected from production and injection wells can be used to infer subsurface flow information during the reservoir development process, which is why effective data-driven technology for general reservoir models is extremely beneficial for reservoir development, the petroleum industry, and geosciences. The injection and production data gathered from wells are always utilized to invert the reservoir's physical parameters and aid in the construction of trustworthy numerical models. These models are capable of simulating the subsurface flow that occurs during the development process and making suggestions for future development. The "data-driven" refers to that, using history data from wells for history matching to improve the numerical model's forecast accuracy, which often needs a repeated numerical simulation procedure. However, history

matching is a difficult and time-consuming task for the most often employed grid-based numerical technology (Maschio et al., 2010; Liu et al., 2013; Ding et al., 2014; Amirsardari et al., 2016; Peng et al., 2017). This is because simulating a full-scale real oilfield takes a lengthy time, and the number of grid parameters to be inverted might reach millions, resulting in a severe multiplicity of solutions. As a result, studies on reservoir numerical simulation have steadily shifted their attention in recent years to the construction of a highly efficient proxy model with sufficient accuracy. As a common proxy model, the reduced-order model may significantly reduce simulation time by employing orthogonal decomposition and segmented stream projection techniques in comparison to the grid-based simulation results (Van Doren et al., 2006; Cardoso and Durlofsky, 2010; He and Durlofsky, 2014). This type of model, however, continues to rely on grid-based three-

dimensional (3D) geological modeling and simulation, which is time-demanding.

In general, the production performance of wells is highly dependent on the physical features bear the wells. Unlike the reduced-order model, several data-driven approaches provide quick modeling, simulation, and history matching of water-flooding reservoirs without requiring knowledge of petrophysical characteristics or other specialized geological information. Earlier research on data-driven models relied heavily on statistical analysis to determine the connectivity between injection-production pairs (Martinez, 1993; Refunjol, 1996; Heffer et al., 1997; Jansen and Kelkar, 1997; Griffiths, 1998). However, because these models do not take into account reservoir physical features or errors of data observation, their predictive ability for production performance is limited. Since then, Albertoni and Lake (2003), and Dinh and Tiab (2007, 2008) have developed a multiple linear regression model that uses the injection rate as the input signal and the liquid production rate as the output signal to estimate the linear coefficient in order to characterize the connectivity between wells and provide a qualitative understanding of reservoir heterogeneity. Additionally, a nonlinear diffusion filter is used to define the temporal delay between injectors and producers in the multiple linear regression model. The multiple linear regression model, on the other hand, requires highly stringent requirements, such as fixed bottom hole pressure and constant flow state. Yousef et al. (2006) developed the capacitance resistance model by introducing the term of reservoir compressibility. capacitance resistance model is named after the analysis between fluid flow in reservoir and capacitive resistant effect in circuit. The allocation factor and time constant are used to characterize the connectivity and fluid storage in well pairs. Notably, capacitance resistance model is capable of efficiently matching history data of the cumulative production rate and bottom hole pressure. Additionally, capacitance resistance model may be used to simulate a well group that has been shut down for an extended period of time. capacitance resistance model is extremely efficient due to the analytic calculation procedure, and several researchers have achieved significant improvements in recent years (Al-Yousef, 2006; Lake et al., 2007; Sayarpour, 2008; Nguyen, 2012; Cao et al., 2015; Naudomsup and Lake, 2019). Nonetheless, this model has certain theoretical limitations, as it frequently attempts history matching on a steady production era, which limits its accuracy (Naudomsup and Lake, 2019).

A novel data-driven model for production optimization was proposed by Lerlertpakdee et al. (2014). The well system in the reservoir is represented by a linked network model in which each pair of wells is connected to a one-dimensional (1D) finite difference reservoir simulation. It is worth noting at this point that the 1D connection element between wells is formed of grids of varying widths and permeabilities. Grid pressure and saturation are semi-implicitly solved. In general, the number of inversed model parameters is a multiple of the number of grids. When a whole-scale reservoir is used, a large number of connection elements results in an increase in the number of possible solutions for history matching and has a negative effect on calculation performance.

Zhao et al. (2015) suggested an inter-well numerical simulation model (INSIM) for water flooding reservoirs by using the advantages of Lerlertpakdee's model (Lerlertpakdee et al., 2014). INSIM simplifies the 3D flow of a grid-based full-scale simulator by solving a sequence of 1D interconnected networks between wells semi-analytically. Two distinctive factors are provided to characterize each connection element in connected networks: transmissibility and control PV. The explicit saturation scheme is utilized to determine the average pressure in the control volume at each well node, and the Buckley-Leverett theory is employed to solve the saturation profile along each connection element. Because the model parameters are substantially smaller than the model parameters of Lerlertpakdee et al. (2014), INSIM simulates more quickly. By absorbing history data, the INSIM is capable of performing rapid history matching and accurate production forecasts.

INSIM family methods can be utilized for effective history matching, production optimization, and connectivity characterization of water-drive reservoirs through continuous development (Guo and Reynolds, 2019; Zhao et al., 2019, 2020). INSIM front tracking (INSIM-FT), for example, developed a front-tracking method for solving the saturation equation, resulting in a more accurate calculation of water cut. By considering fluid gravity, INSIM-FT in three-dimensions with gravity (INSIM-FT-3D) can simulate and calculate the flow of an arbitrary route between distinct perforating locations in the 3D reservoir. The novel Riemann solver is used, which is based on a convex-hull approach and allows for the solution of the saturation equation with gravity. To improve the characterization of reservoir physical features, enrich the fluid flow path, and enhance the capability of history matching, INSIM-FT and INSIM-FT-3D propose to add virtual nodes to the INSIM (i.e., the liquid rate is zero). However, because the addition of virtual nodes obscures and complicates the connectivity between actual well nodes, it is vital to investigate how to explore the connectivity between real well nodes in the presence of virtual nodes. To do this, interwell numerical simulation model with flow-path tracking (INSIM-FPT) offers a path tracking approach and a method for refining virtual well nodes that is distinct from INSIM-FT in order to optimize the INSIM model. INSIM-FPT is capable of rapidly matching the history of a real reservoir and accurately forecasting its production performance, as well as dynamically revealing changes in inter-well control PV and transmissibility in the presence of a high number of virtual well nodes.

Borregales et al. (2020) and Ren et al. (2019) recently proposed mapping the flow network to a two-dimensional (2D) simulation grid, in which the connections in the flow system are still more complex than the 1D connection element used in INSIM family methods or the 1D difference grid used by Lerlertpakdee et al. (2014), and thus the complexity of history matching is increased. Kiær et al. (2020) introduced FlowNet, which retains the flow network's 1D difference grid, and applied it to the Brugge benchmark model. However, there is insufficient definition of the connection elements' characteristic parameters in FlowNet, as well as the constraint on the connection elements' control volumes. Additionally, the performance of the FlowNet and INSIM family methods is not

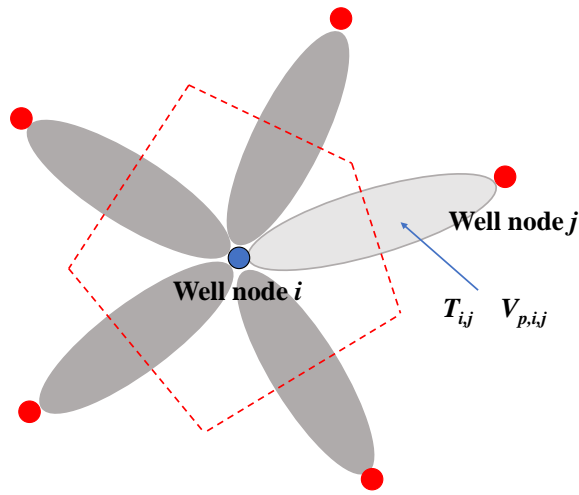


Fig. 1. 1D connection elements in the simplified injection-production system of the reservoir.

compared, nor is the study of earlier methods based on flow networks (i.e., INSIM family methods and the model from Lerlertpakdee et al. (2014) analyzed.

In comparison to pure data-driven models (Wood and Choubineh, 2019; Wu et al., 2020), these physics-based data-driven models can incorporate physical information about the flow equation, hence improving the model's prediction capacity. To facilitate the construction of a more general physics-based data-driven flow-network model and the clarification of the differences and similarities from these models (particularly for the INSIM family methods), this paper proposes a general physics-based data-driven framework (GPDF). GPDF inherits the concept of INSIM family methods and combines them with grid-based approaches for reservoir simulation and history matching. Thus, the GPDF presented in this paper not only benefits from the INSIM family methods, which significantly reduce model order to increase calculation efficiency and facilitate history matching, but also from the traditional grid-based method, which has a high degree of generalization for the flow models. Additionally, the ideas contained in GPDF may have a great deal of promise for widespread use in the disciplines of science and engineering.

2. Methodology

2.1 A brief review of INSIM family methods

Original INSIM. The primary objective of the original INSIM model is to reduce the reservoir's injection-production system to a network consisting of a sequence of 1D connection elements. As illustrated in Fig. 1, The blue or red solid circle denotes the well node, the red dotted polygon denotes the control volume of well node i , and the gray ellipse denotes the one-dimensional connection element between two well nodes. Two factors are used to characterize each connection element: transmissibility $T_{i,j}$ and control PV $V_{p,i,j}$. The transmissibility indicates the fluid's ability to flow through the connection element, whereas the control PV indicates the connection element's pore volume. Transmissibility is a term that refers

to the permeability of the connection element, whereas the control PV is a term that refers to the cross-sectional area of the connection element.

INSIM's original implementation used a sequentially coupled approach for the flow and transport equations, and the calculation procedure is primarily comprised of three steps: (1) implicit calculation of the average pressure in the well node's control volume (relative permeability in mobility is determined by the last time step's saturation); (2) estimation of the upstream flux on each 1D connection element using the calculated average pressure in the well node's control volume; (3) calculation of the water saturation on each 1D connection element. Following that, the following specifics of the three steps are briefly described:

Step 1: pressure calculation based on mass-balance equations

The original INSIM method is based on isothermal two-phase oil-water flow. The material balance equation is built by considering the control volume $V_{p,i}$ of the well node i as the object.

$$\sum_{j=1}^{n_{c,i}} T_{i,j}(p_j(t) - p_i(t)) + q_i(t) = \frac{dp_i(t)}{dt} c_{t,i}(t) V_{p,i}(t) \quad (1)$$

where t is time; $n_{c,i}$ denotes the number of wells connected to well node i ; $q_i(t)$ means the source and sink terms (i.e., production or injection) of well node i , m^3/day . If well node i is a production well, the term is positive. If well node i is a water injection well, the term is negative; p_i is the average pressure in the control volume $V_{p,i}$; $c_{t,i}$ is total compressibility in the control volume of the well i . The original INSIM method solves the average pressure of the well-node control volume using the implicit pressure and explicit saturation scheme (IMPES), that is

$$p_i^n - p_i^{n-1} = \frac{\Delta t^n}{c_{t,i}^{n-1} V_{p,i}^{n-1}} \left[-p_i^n \sum_{j=1}^{n_{c,i}} T_{i,j}^{n-1} + \sum_{j=1}^{n_{c,i}} T_{i,j}^{n-1} p_i^n + q_i^{n-1} \right] \quad (2)$$

where the superscript n and $n-1$ denote the n and $n-1$ th time step, respectively. And

$$T_{i,j}^{n-1} = T_{i,j}^0 \frac{\lambda_{ij}^{n-1}}{\lambda_{ij}^0} = \frac{k_{ij} A_{ij} \lambda_{ij}^{n-1}}{L_{ij}} \quad (3)$$

$$V_{p,i}^{n-1} = V_{p,i}^0 [1 + C_{t,i}(p_i^{n-1} - p_i^0)] \quad (4)$$

$$C_{t,i}^{n-1} = C_r + S_{w,i}^{n-1} C_w + S_{o,i}^{n-1} C_o \quad (5)$$

where A_{ij} and L_{ij} represent the flow cross-sectional area and the length of connection element between well i and well j , respectively; $T_{i,j}^0$ and $T_{i,j}^{n-1}$ are respectively the transmissibility values between well i and well j at the initial time and $n-1$ th time step; $V_{p,i}^0$ and $V_{p,i}^{n-1}$ denote the control volume of well node i at the initial time and $n-1$ time step, respectively; $\lambda_{i,j}^0$ and $\lambda_{i,j}^{n-1}$ successively refer to the mobility values between well i and well j at the initial time and $n-1$ time step; C_r , C_w and C_o successively refer to the compressibility coefficients

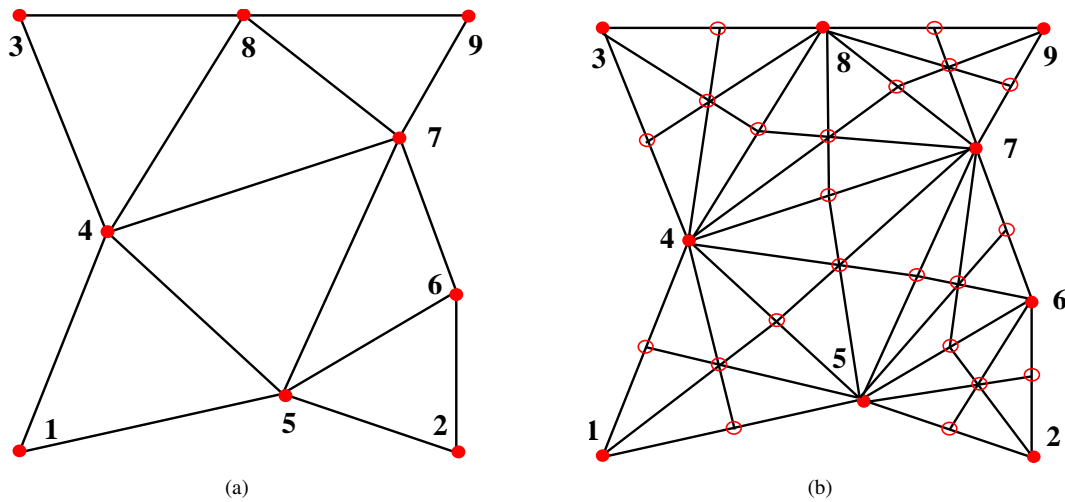


Fig. 2. Sketch of addition of virtual nodes. (a) connection map constructed by real well nodes and (b) refined connection map with virtual nodes.

of reservoir rocks, water and oil; $S_{w,i}$ and $S_{o,i}$ are the water saturation and oil saturation at well node i , respectively.

Step 2: approximation of upstream flux

If the average pressure of the control volume of well node i is larger than that of well node j on a connection element, well node i might be considered to be upstream of the connection element. To solve for saturation of the 1D connection element, the flow flux between wells i and j is approximated at the n^{th} time step as

$$q_{i,j}^n = T_{i,j}^{n-1} (p_i^n - p_j^n) \quad (6)$$

Step 3: saturation calculation based on Buckley-Leverett theory

The original INSIM uses the Buckley Leverett theory to calculate oil-water two-phase flow on a 1D connection element. By disregarding capillary force, gravity, and compressibility, the pure-convection transport equation for water saturation is obtained.

$$\frac{\partial S_w(s,t)}{\partial t} + \frac{qL_{i,j}}{V_{p,ij}} \frac{\partial f_w(x,t)}{\partial x} = 0 \quad (7)$$

where x represents the distance from a certain point to the injection end, and f_w denotes the water cut at x ; η is a constant. By solving the pure-convection transport equation using the characteristic line approach, the following equation for the movement of an equal-saturation surface can be obtained

$$x - x_0 = \frac{f'_w(S_w)}{\phi A} \int_0^t Q dt \quad (8)$$

where Q and $\int_0^t Q dt$ are the water injection rate and cumulative water injection quantity, respectively. The original INSIM can obtain the water injection rate of each time step and the cumulative injected water on the 1D connection element via Eq. (6). Therefore, the water saturation profile on the connection element can be calculated by Eq. (8).

INSIM-FT. Guo et al. (2018) solved Eq. (7) using a front tracking method to produce a more accurate prediction of water cut. This method decomposes the Cauchy problem into a

set of Riemann problems with analytic solutions. Allowing for the length L_{ij} of the connection element between the upstream and downstream well nodes, the water saturation at time t can be stated as

$$S_{w(x,t)} = \begin{cases} S_{w,i}, & \frac{x-x_0}{t} < \eta f'_w(S_{w,i}) \\ (f'_w)^{-1} \frac{x-x_0}{\eta t}, & \eta f'_w(S_{w,i}) < \frac{x-x_0}{t} < \eta f'_w(S_{w,j}) \\ S_{w,j}, & \frac{x-x_0}{t} < \eta f'_w(S_{w,j}) \end{cases} \quad (9)$$

where x_0 is the location of the single discontinuity in the initial condition. To improve the description of reservoir properties and to enrich the flow path between injector and producer, INSIM-FT proposed for the first time the addition of virtual nodes to the INSIM.

INSIM-FT-3D. Guo extended the INSIM-FT model from a 2D to a 3D reservoir. By incorporating gravity into the material balance equation, it is possible to simulate the flow of various perforating locations and any well route between wells. Notably, INSIM-FT-3D implemented a new Riemann solver based on a convex-hull approach that enables the Buckley-Leverett problem to be solved with gravity. As illustrated in Fig. 2, INSIM-FT-3D adopted Delaunay triangulation to create a 3D connection map of the reservoir, which more clearly depicts the flow interaction between many perforated wells in different levels. The subsequent INSIM-FPT employed a different method of triangulation, although its core idea is similar to that of INSIM-FT-3D.

INSIM-FPT. In general, increasing the number of virtual well nodes obscures and impairs the transmissibility between real well nodes. As a result, it is required to investigate the interaction of real well nodes with virtual well nodes. INSIM-FPT optimized the INSIM by introducing a path tracking technique and a variety of infilling methods for virtual well nodes based on INSIM-FT and INSIM-FT-3D. On the one hand, INSIM-FT-3D is capable of efficiently matching history data and forecasting production. On the other hand,

it can reveal dynamic changes in inter-well control PV and transmissibility when there are a high number of virtual well nodes. After calculating the pressure at the well nodes, the inter-well network transforms into a directed graph, in which the direction of each connection element is defined as from the high-pressure well node to low-pressure well node. Then, using graph theory's path tracking algorithm, all paths between any two well nodes may be determined.

The direct allocation factor d_{af} between well node i and j is calculated by

$$d_{af} = \frac{q_{i,j}^n}{\sum_{k=1}^{N_j} q_{k,j}^n} \quad (10)$$

where N_j denotes the number of well nodes connected to well node j with a pressure less than that of well node j ; $\sum_{k=1}^{N_j} q_{k,j}^n$ and denotes the total flow rates exiting well node j .

The dynamic control volume (V_c) for flow paths that connects well node i and j is given by

$$V_{c,i,j}^n(s) = V_{p,i,b_1}^n \frac{Q_{i,j}^n(s)}{q_{i,b_1}^n} + V_{p,b_1,b_2}^n \frac{Q_{i,j}^n(s)}{q_{b_1,b_2}^n} + \dots + V_{p,b_{int},b_j}^n \frac{Q_{i,j}^n(s)}{q_{b_{int},b_j}^n} \quad (11)$$

where b_1, b_2, \dots, b_{int} denote the intermediate virtual well nodes in the flow path and the subscript *int* is the number of intermediate virtual well nodes along the flow path; $Q_{i,j}^n(s)$ is the corresponding allocated flow rate to the flow path.

2.2 Analysis of the limitations of INSIM family methods

After a brief overview of the INSIM family methods in Section 2.1, we can summarize the INSIM family methods' core concepts as follows:

(1) Simplified reservoir model characterization: the reservoir flow is simplified to be the flow through a connected network made of 1D connection elements. This concept reduces the complex multi-dimensional flow to a sequence of 1D flow, and the 1D connection element is only defined by transmissibility and control PV, implying that the 1D connection element is of equal width and permeability, in contrast to Lerlertpakdee et al. (2014), who present an unequal-width 1D connection element. Thus, the geological parameters of the reservoir can be reduced from a large number of grid parameters (permeability, porosity, etc.) in the conventional grid-based numerical simulator to the characteristic parameters (i.e., transmissibility and control PV) of a small number of 1D connection elements, thereby increasing the efficiency of history matching and easily achieving the accuracy required by engineering at the reservoir scale;

(2) Calculation of two-phase flow equations quickly along a 1D connection element: The INSIM family methods begin by calculating the average pressures of well-node control volumes, and then using the average pressure gradient to estimate the upstream flux on a 1D connection element (because the size of the well-node control pore volume is generally large, this estimation is relatively rough, as described in detail later in this section), and finally using Buckley-Leverett theory or

the Riemann method to calculate the water saturation along the connection element analytically. Calculations on the 1D connection element can also be performed quickly using the finite volume/difference approach or other numerical methods, as the computational cost of numerical methods is typically small for 1D flow problems. For instance, the upwind finite difference method can be used to quickly compute the pure-convection equation satisfied by water saturation. In comparison to the Buckley-Leverett theory or the Riemann approach based on analytical theory, the numerical method for 1D flow can swiftly solve a broader variety of flow problems, except for pure convective water saturation equations in two-phase flow.

(3) Tracking flow paths and establishing an inter-well injection-production relationship: to some extent, the algorithm for establishing an inter-well injection-production relationship based on fast path searching is a direct inference that reduces the reservoir model to a connected network composed of 1D connection elements. As a result, this property is nearly identical to the first in the INSIM family method.

The above analysis of the INSIM family methods' essential characteristics reveals that the simplified reservoir characterization reflected in the first and third characteristics is critical to their success, while the calculation method for the reservoir fluid's unknown quantities (pressure, saturation, etc.) can be changed in the second.

Additionally, this study presents GPDPF based on the results of the preceding investigation. Before formally presenting GPDPF, this paper discusses some of the constraints associated with the INSIM-family methods due to the second characteristic listed above.

The flow model's generalizability is limited. The computation of water saturation on the 1D connection element in the original INSIM is based on the Buckley-Leverett theory (Zhao et al., 2016), which has a high analytical accuracy and calculation efficiency. Guo et al. calculated the oil-water two-phase flow using a Riemannian problem solver. However, when considering other flow problems, such as three-phase flow, steam flooding, chemical flooding, or other compositional models, these analytical or semi-analytical methods are ineffective, and when solving the equation with diffusion term on the 1D connection element, it is necessary to determine both the upstream and downstream flux. As a result, numerical approaches are still required to accurately handle complex flow issues on 1D connection elements. Additionally, if the flow model is complicated, the information included in the upstream or downstream flux will be complex, requiring extensive and time-consuming labor to adapt the INSIM family methods to the complex flow model. Naturally, the INSIM family methods may employ parallel computing to solve equations on 1D connection elements, as each connection element's calculations are independent.

For example, when non-isothermal flow is considered, the heat conduction and convection equations (including the diffusion term) on a 1D connection element must be calculated. When adequate accuracy is desired, a numerical method is frequently required. The finite volume method (FVM) is frequently used to calculate the coupled heat and mass transfer.

Table 1. Computed upstream flux at the different times via Eq. (6).

Time (day)	Control-volume pressure gradient of well nodes (MPa/m)	Water saturation within the leftmost grid (fraction)	The upstream total mobility (mPa·s) ⁻¹	Computed upstream flux (m ³ /day)
5	0.0046	0.1538	0.00846	1.34
10	0.0087	0.1571	0.00885	2.66
15	0.0123	0.1602	0.00922	3.92
20	0.0155	0.1632	0.00958	5.13
25	0.0183	0.1659	0.00991	6.27
30	0.0207	0.1686	0.01023	7.32
35	0.0228	0.1710	0.01052	8.29
40	0.0246	0.1733	0.01080	9.18
45	0.0261	0.1756	0.01107	9.99
50	0.0274	0.1777	0.01132	10.72

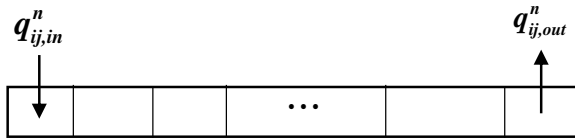


Fig. 3. Sketch of the divided grid in one 1D connection element.

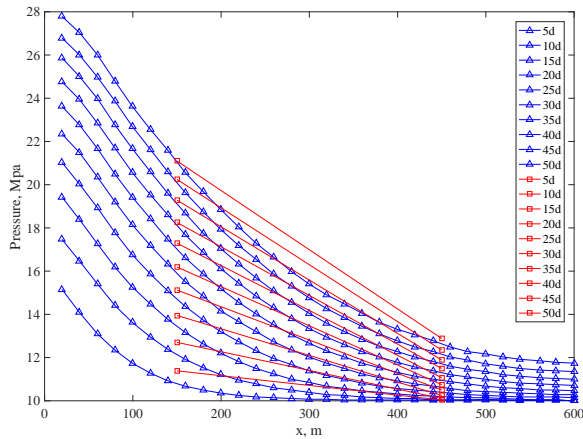


Fig. 4. Pressure profiles on a 1D connection element at different times.

As illustrated in Fig. 3, it is important to calculate the upstream and downstream fluxes on the 1D connection element in order to solve for the saturation and temperature distributions on the 1D connection element. If we continue to employ the INSIM family methods, significant changes to the algorithm framework are required, and precisely estimating the downstream flux on the 1D connection element is a difficult task.

In Eq. (6), using the gradient of the average pressure of the well-node control PV to estimate the upstream flux is insufficiently accurate, even more so when the viscosity difference between displaced and flooding fluids is significant. As an illustration of oil-water flow, consider well node *i* as the injector and well node *j* as the producer. The viscosity of the oil is 100 mPa·s, while the viscosity of the water is

1 mPa·s. When water is injected because the oil viscosity is larger than that of water, it is reasonable to assume that the injected water will concentrate near well *i* resulting in a pressure gradient near well *i* that is significantly bigger than the well-node control volume. As illustrated in Fig. 4, the blue curve depicts the pressure distribution at various times, whereas the red line reflects the control volume’s average pressure distribution at various times. At the early stages of water injection, the pressure gradient near the left injection well will be much larger than the average pressure gradient of the control volume. If the flow rate associated with the pressure gradient between the first and second grids is assumed to be a constant 10 m³/day (it will be less than 10 m³/day, as the first grid will provide elastic energy to store a certain amount of injected water), the water injection rate on the 1D connection element estimated by Eq. (6) can be calculated in Table 1. As can be seen from Eq. (6), the water injection rate is significantly less than the actual water injection rate of 10 m³/day in the beginning, and it does not reach that level until 50 days.

2.3 GPDF

GPDF retains the essence of the INSIM family methods, in that the reservoir model is simplified to a connected network made of 1D connection elements, and the characteristic parameters of each connection element are determined through data-driven optimization (history matching). It is, however, distinct from the INSIM family methods, which compute the average pressure of the control volume of well nodes first to estimate the upstream flux and then compute the water saturation on the connection element. In GPDF, the grids are divided on the connection element (it should be noted that the grid on the same connection element must have an equal flow cross-sectional area and permeability), and the pressure, saturation, concentration in compositional models, and other physical quantities are accurately calculated using the fully implicit scheme of flow equations and Newton iteration based nonlinear solver that are commonly used in the traditional grid-based reservoir numerical simulator.

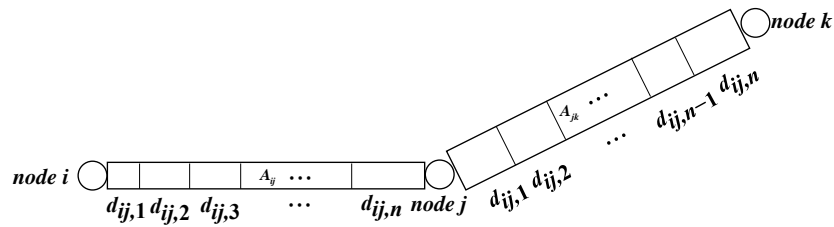


Fig. 5. Sketch of the grid divided into the connection elements in GPDF.

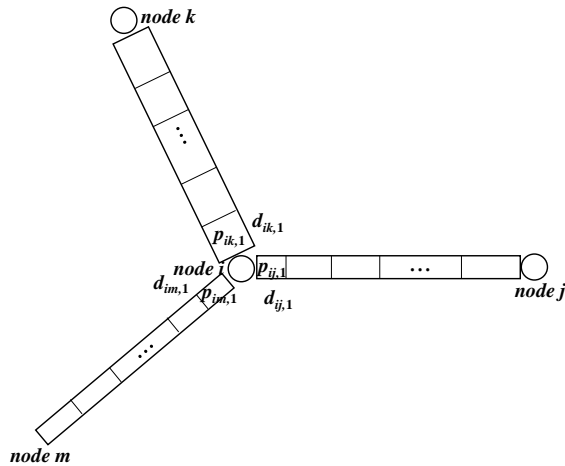


Fig. 6. Three connection elements which are connected with a well node *i*.

Reservoir model and computation of governing equations. The basic assumption and treatment of GPDF are as follows:

Before introducing the details of GPDF, it should be noted that this work characterizes the connection element using the volume of the connection element, not the pore volume. This is because the porosity of the connection element changes when pressure changes, whereas the volume of the connection element remains constant. As a result, the total volume of connection elements should be controlled rather than the total pore volume throughout the history matching (data-driven) phase, as the total pore volume may be bigger or smaller than the initial reservoir pore volume due to compressibility.

Similarly to the INSIM family methods, GPDF creates a well-to-well connected network by utilizing well nodes (water injection well, production well) and virtual well nodes. Using the 2D model as an example, see Fig. 5. Node *i* represents a water injection well, node *j* represents a virtual well, and node *k* represents a production well. However, unlike the INSIM family methods, some grids in GPDF are separated into individual connection elements for numerical methods using a fully implicit scheme.

The 1D connection element has an equal cross-sectional area and inherits the properties of the INSIM family methods. In contrast to Lerlertpakdee et al. (2014), the cross-sectional area is determined by the ratio of the control volume (not the pore volume) to the distance between well nodes, or

$$A_{ij} = \frac{V_{ij}}{L_{ij}}, \quad A_{jk} = \frac{V_{jk}}{L_{jk}} \quad (12)$$

where V_{ij} and V_{jk} are volumes of connection element *i-j* and *j-k*.

The permeability of the 1D connection element is homogeneous, which is calculated from the transmissibility and control PV of the connection element

$$k_{ij} = \frac{T_{ij}L_{ij}^2}{V_{ij}}, \quad k_{jk} = \frac{T_{jk}L_{jk}^2}{V_{jk}} \quad (13)$$

The 1D connection element is divided into some grids, which might be uniform in length or non-uniform in length. Eq. (14) calculates the transmissibility between grids using the harmonic average scheme. Assuming that *a* and *b* are two adjacent grids on the connection element *ij*, the transmissibility between *a* and *b* is as follows

$$T_{ij,ab} = (T_{ij,a}^{-1} + T_{ij,b}^{-1})^{-1}, \quad T_{ij,a} = \frac{k_{ij}A_{ij}}{d_{ij,a}/2}, \quad (14)$$

$$T_{ij,b} = \frac{k_{ij}A_{ij}}{d_{ij,b}/2}$$

Each 1D connection element is connected to the grid on which the well is located, forming an inter-well connected network composed of a 1D grid. As illustrated in Fig. 6, the grid on which the well is located is circular, with a radius comparable to the equivalent radius in a conventional reservoir model, and the radius and grid size can be adjusted to meet accuracy requirements. It is worth noting that the permeability of the grid on which the well is located is independent of the permeability of the well node and has nothing to do with the permeability of the 1D connection element that connects the well node. Assuming that the grid *c* on the connection element *ij* is connected to the well node *i* the transmissibility between the well node *i* and grid *c* is determined as follows.

$$T_{ij,ic} = (T_{ij,i}^{-1} + T_{ij,c}^{-1})^{-1}, \quad T_{ij,i} = \frac{k_{ij}A_{ij}}{r_w}, \quad T_{ij,c} = \frac{k_{ij}A_{ij}}{d_{ij,c}/2} \quad (15)$$

The governing equation is discretized using FVM on the basis of the developed inter-well connection network, and its fully implicit scheme is constructed to simultaneously solve for pressure, saturation, or other physical parameters. Using two-phase oil-water flow as an example, the following is the fully implicit discretization scheme for the flow equations

$$-\sum_{j=1}^n [\lambda_{o,ij} T_{ij} (p_{o,i} - p_{o,j})]^{t+\Delta t} + q_{osc}^{t+\Delta t} = \frac{\Delta V_i}{\Delta t} \left[\left(\frac{\phi_i S_{o,i}}{B_{o,i}} \right)^{t+\Delta t} - \left(\frac{\phi_i S_{o,i}}{B_{o,i}} \right)^t \right] \quad (16)$$

$$\begin{aligned}
& - \sum_{j=1}^n [\lambda_{w,ij} T_{ij} (p_{o,i} - p_{o,j} + p_{cwo,j} - p_{cwo,i})]^{t+\Delta t} + q_{wsc}^{t+\Delta t} \\
& = \frac{\Delta V_i}{\Delta t} \left[\left(\frac{\phi_i S_{w,i}}{B_{w,i}} \right)^{t+\Delta t} - \left(\frac{\phi_i S_{w,i}}{B_{w,i}} \right)^t \right]
\end{aligned} \quad (17)$$

where T_{ij} and $\lambda_{a,ij}$ are the transmissibility and phase mobility of the connection.

T_{ij} is the half of the harmonic mean of two half-transmissibility, that is

$$T_{ij} = \frac{T_i T_j}{T_i + T_j} \quad (18)$$

$\lambda_{a,ij}$ is expressed as

$$\lambda_{a,ij} = \frac{k_{ra,ij}}{\mu_{a,ij} B_{a,ij}} \quad (19)$$

and the upstream scheme and arithmetic average scheme are used for the terms (relative permeability) subject to the saturation and the terms (viscosity and volume factor) subject to pressure, respectively, these are

$$\begin{aligned}
\mu_{a,ij} &= \frac{\mu_{a,i} + \mu_{a,j}}{2}, \quad B_{a,ij} = \frac{B_{a,i} + B_{a,j}}{2} \\
k_{ro,ij} &= \begin{cases} k_{ra,i} & \text{if } P_{a,i} \geq P_{a,j} \\ k_{ra,j} & \text{if } P_{a,i} < P_{a,j} \end{cases}
\end{aligned} \quad (20)$$

The nonlinear solver based on Newton iteration and automatic differentiation method can be used to solve the above-mentioned fully-implicit discrete equations, so as to obtain the pressure and saturation of each grid. It can be known that if GPDF is applied to the flow problems other than oil-water two-phase flow, it is only necessary to replace the discrete equations in Eqs. (16) and (17) with the fully-implicit finite-volume discrete scheme of the new flow governing equations, and then solve them in the same way. Therefore, compared with the INSIM family method, the GPDF not only has better generalization of flow models but also has higher computational accuracy in theory for the calculation of the upstream and downstream fluxes of 1D connection elements.

Data-driven (history matching) procedure. We perform history matching on the reservoir's production data and automatically update the characteristic characteristics of the 1D connection elements based on the proposed GPDF. Our objective with history matching is to minimize the discrepancy between predicted and actual values. As defined above, the objective function is as follows

$$O(m) = \frac{1}{2} [g(m) - d_{obs}]^T C_D^{-1} [g(m) - d_{obs}] \quad (21)$$

where m represents the model parameters, including the transmissibility and control volume, $g(m)$ represents the simulation of GPDF, and d_{obs} represent the actual production data. Simultaneous Perturbation Stochastic Approximation (SPSA) algorithm is used to solve the optimization problem in history matching (Cao et al., 2014, 2017). The control pore volume of each connection element is modified continually throughout

SPSA iterations of INSIM family methods, but the sum of them is often set equal to the overall reservoir volume (Zhao et al., 2016). As discussed above, the cumulative pore volume of all connection elements may be greater or less than the initial reservoir pore volume due to compressibility. Thus, volume is utilized in place of the connection element's pore volume to characterize it, and the sum of the connection element's control volumes should be less than or equal to the reservoir's total volume. The total volume of connection components after history matching is equivalent to the impacted region of injection-production displacement. When a zone of high permeability exists between injection and production, the total volume after history matching may be near to the volume of the high-permeability zone rather than the reservoir's complete volume. The range of the zone of high permeability is similar to the active displacement area. As a result, this paper proposes that the constraint on the sum of volumes of connection elements in the data-driven (i.e., history matching) process in GPDF be an unequal constraint, i.e.

$$0 < V \leq V_{total} \quad (22)$$

where V_{total} is the total volume of the reservoir model and V is the total volume of all connection elements. An example will be given in the next section to prove this issue.

3. Numerical examples

3.1 Validation of the unequal constraint

A test case of water flooding channeled reservoir is used to demonstrate the validity of unequal constraint in this example. As illustrated in Fig. 7, the high-permeability band has a permeability of 500 mD, while the remaining regions have a permeability of 0.01 mD. One injector is located in the middle reservoir, and four producers are located in each of the reservoir's four corners. Table 2 summarizes the reservoir physical properties and well controls that are relevant. As illustrated in Fig. 7(b), we create the GPDF model using the reservoir's fundamental physical parameters. Eclipse 2011 is used to calculate 100 days of oil production rate data, with

Table 2. Physical properties used in example 1.

Properties	Values
Porosity	0.3
Oil volume factor	1.2
Grid size	10 m
Water viscosity	0.6
Rock compressibility	1.07e-4 MPa ⁻¹
Water volume factor	1.000
Oil compressibility	3.02e-3 MPa ⁻¹
Oil density	820 kg/m ³
Water compressibility	5e-4 MPa ⁻¹
Water density	1000 kg/m ³
Oil viscosity	2 cp
Initial oil saturation	0.85

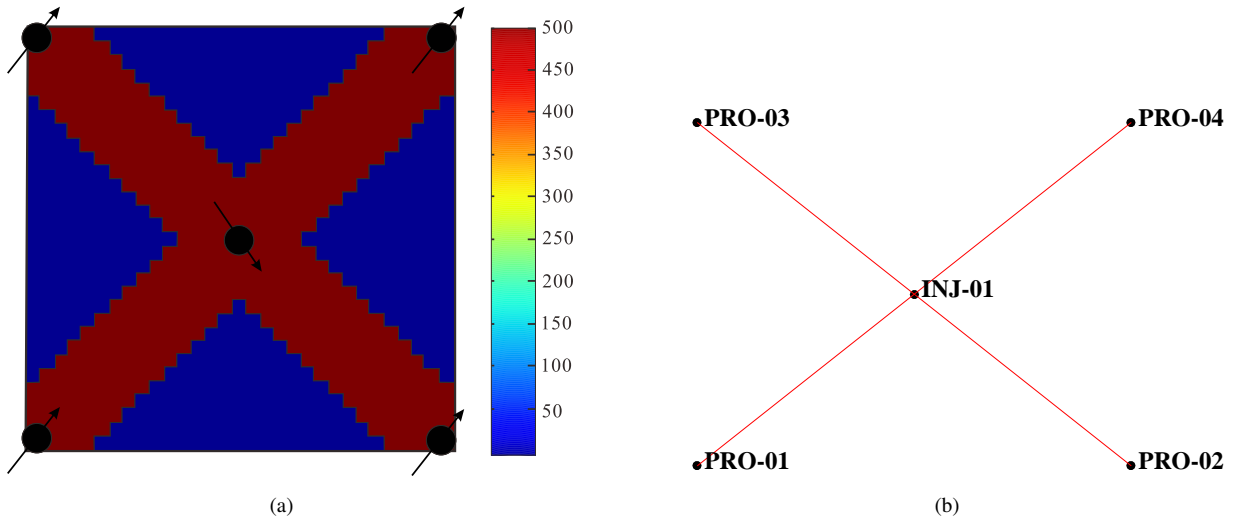


Fig. 7. The reservoir model and its data-driven network model over (a) reservoir model based on Cartesian grid and (b) reservoir model based on connection elements.

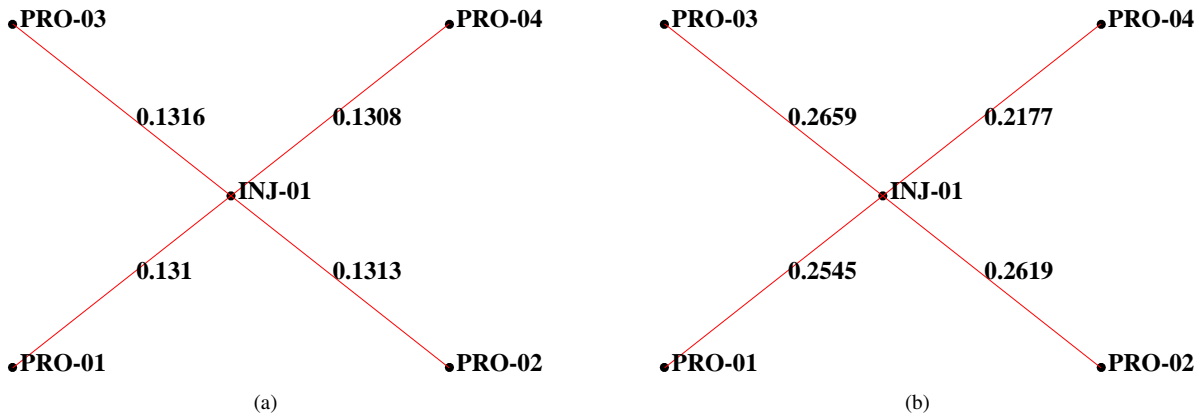


Fig. 8. Comparison of the ratio of control volumes to total reservoir volume over (a) unequal constrain in GPDF and (b) equal constrain.

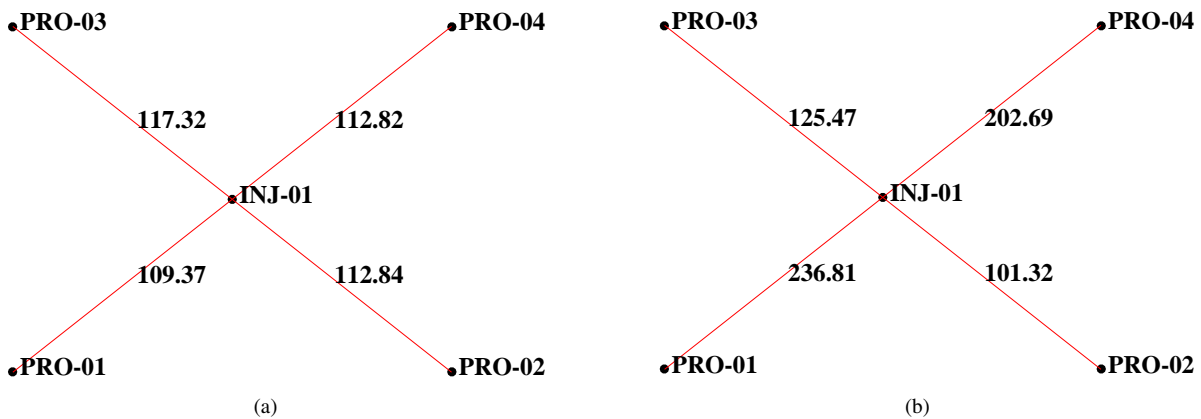


Fig. 9. Comparison of the transmissibility of connection elements over (a) unequal constrain in GPDF and (b) equal constrain.

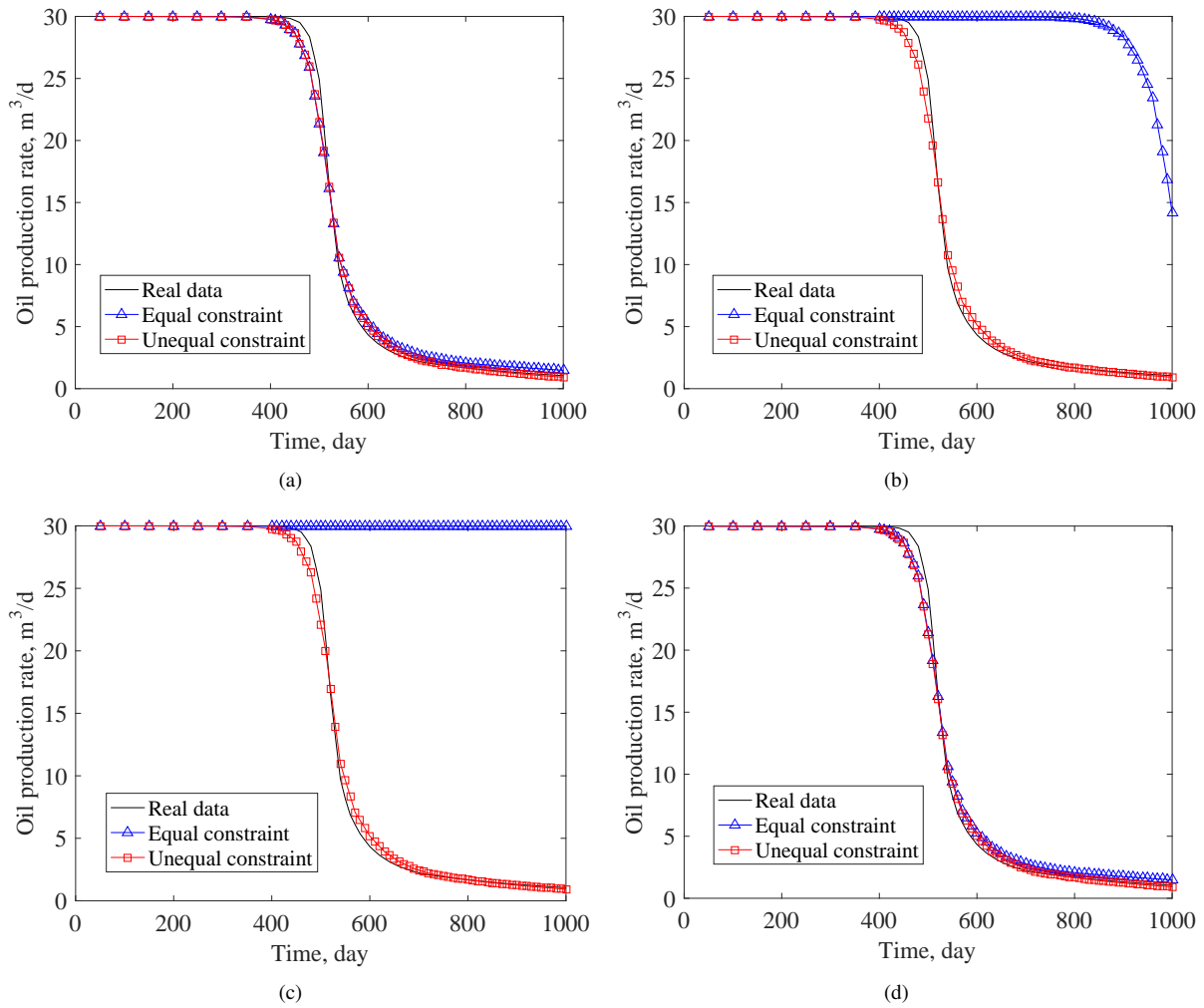


Fig. 10. History-matched single-well oil production rates of example 1 over (a) Well #1, (b) Well #2, (c) Well #3 and (d) Well #4.

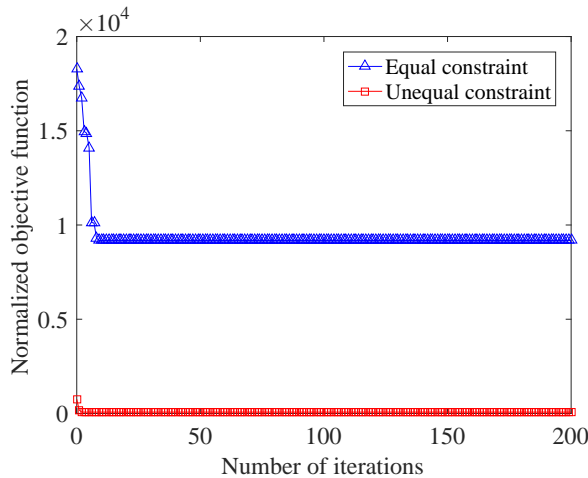


Fig. 11. Normalized objective function vs. number of iterations of example 1.

the first 60 days serving as the history data to match. The history matching technique is performed in 200 iteration steps using the equal and unequal constraints, respectively. Figs. 8-11 plot the inverted model parameters and the normalized

target function value against the number of iterations. As can be observed, the unequal constraint achieves a higher prediction of performance on wells 2 and 3 than the equal constraint does after 200 iteration steps. Furthermore, because the low permeability section has a permeability of only 0.01 mD, the reservoir model’s real flow area is virtually entirely comprised of the high-permeability channeled portion. Notably, the volume coefficient of the connection element derived from the unequal constraint is approximately 0.1310, which is commensurate with the fraction of channeled volume in the total reservoir volume. However, when equal restriction is applied, the control volume of the connection elements PRO-01-INJ-01 and PRO-04-INJ-01 is nearly twice that of the high permeability area, which is unphysical. As a result, the unequal constraint in GPDF is more physical than the equal constraint utilized in the INSIM family methods.

3.2 Better performances of GPDF than INSIM-family methods

As mentioned in Section 2, the proposed GPDF’s pressure, saturation, and other physical quantities are simultaneously calculated using the fully implicit scheme. In theory, GPDF is

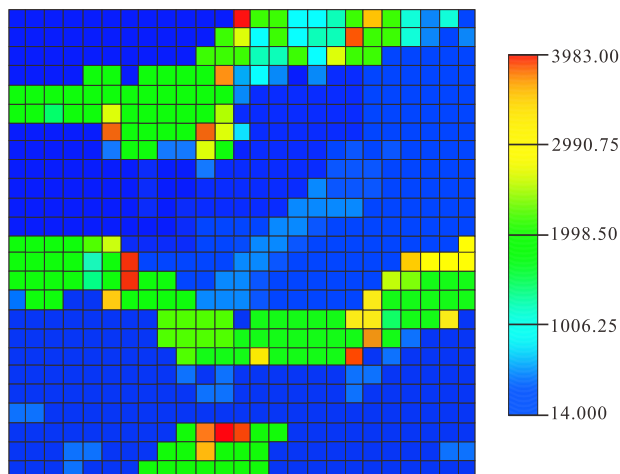


Fig. 12. Permeability distribution of reservoir model.

Table 3. Pore fractal dimensions based on FHH model.

Properties	Values
Porosity	0.2
Oil volume factor	1.01
Grid size	10 m
Water viscosity	0.5
Rock compressibility	$8.84e-3 \text{ MPa}^{-1}$
Water volume factor	1.0042
Oil compressibility	$1.45e-3 \text{ MPa}^{-1}$
Oil density	912 kg/m^3
Water compressibility	$5.8 e-4 \text{ MPa}^{-1}$
Water density	999 kg/m^3
Oil viscosity	20 cp
Initial oil saturation	0.8

more accurate than the INSIM family methods at estimating the connection element's upstream flux. This section will demonstrate how to validate this inference using a heterogeneous reservoir with two-phase flow. As illustrated in Fig. 12, this heterogeneous reservoir contains four producers and five injectors. Table 3 summarizes the reservoir properties and well controls that are relevant. The Eclipse 2011 commercial simulator is used to model the oil production rate for 100 days in order to obtain the data for the oil production rate. The first 60 days are utilized to match historical data, while the latter 40 days are used to forecast production. The connectivity model is created in Fig. 13. Each connection unit in GPDF is separated into five grids. Then, in 50 iteration steps, we do history matching on INSIM-FPT and GPDF using the oil production rate data. On the one hand, as illustrated in Fig. 16, the GPDF-based production estimate for well PRO-01 is nearly identical to ECLIPSE and superior to INSIM-FPT, a result that holds true for all production wells. Additionally, it can be observed in Fig. 17 when the value of the regularized objective function is plotted against the number of iterations, that the value of GPDF's objective function is always less than

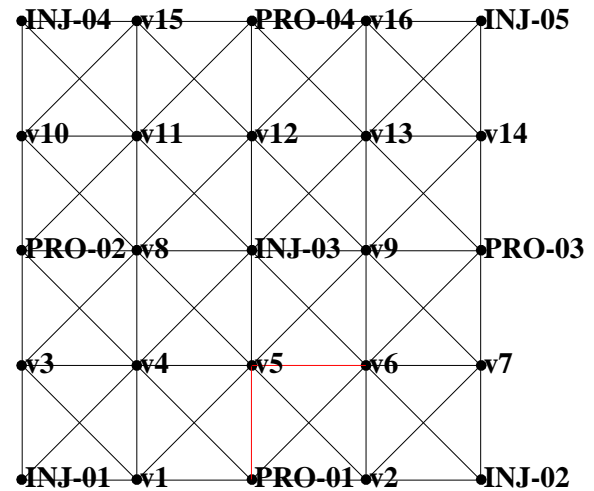


Fig. 13. Reservoir network model.

that of INSIM-FPT. In comparison to INSIM-FPT, the inversed parameters in GPDF are more consistent with the reservoir's physical parameters. For example, in Fig. 14, the connection elements' volume of GPDF is often less than that of INSIM-FPT when unequal constraint is used. Otherwise, the high transmissibility of connection elements in GPDF can correctly reflect the reservoir's high-conductivity channel. As illustrated in the area bounded by the black dotted line in Fig. 15, the area with inversed large transmissibility is extremely close to the reservoir's high-permeability zone. The comparisons above demonstrate that GPDF is more accurate than the INSIM family methods. Additionally, it demonstrates the efficacy of the reservoir network model.

3.3 Validation of good generalization of flow models: Single-phase depletion development

Because the INSIM family methods are limited to two-phase oil-water flow, this section will demonstrate that the given GPDF has a higher degree of generalization than the INSIM family methods. Fig. 18(a) illustrates the permeability distribution of a grid-based reservoir model with five production wells positioned in the reservoir's center and four corners and a fixed bottom hole pressure of 20 MPa. Other physical properties of the reservoir that are necessary are listed in Table 4. In Fig. 18(b), a simple reservoir network model for GPDF is built. The first 60 days of oil production rates for each well are gathered using ECLIPSE, and the first 15 days of data are used as history data. The network model is used to do history matching, and the estimated oil production rates for the following 45 days are calculated using GPDF. Fig. 19 compares the GFDM results to the ECLIPSE data, Fig. 20 depicts the matched transmissibility and control pore volume of each connection element, and Fig. 21 depicts the normalized objective function vs. SPSA iterations. As can be observed, the accuracy of GPDF's history matching and prediction is quite great, and the inversed transmissibility essentially represents the reservoir's high-permeability zone. Due to the fact that the network model in Fig. 18(b) has 28 connection elements, there

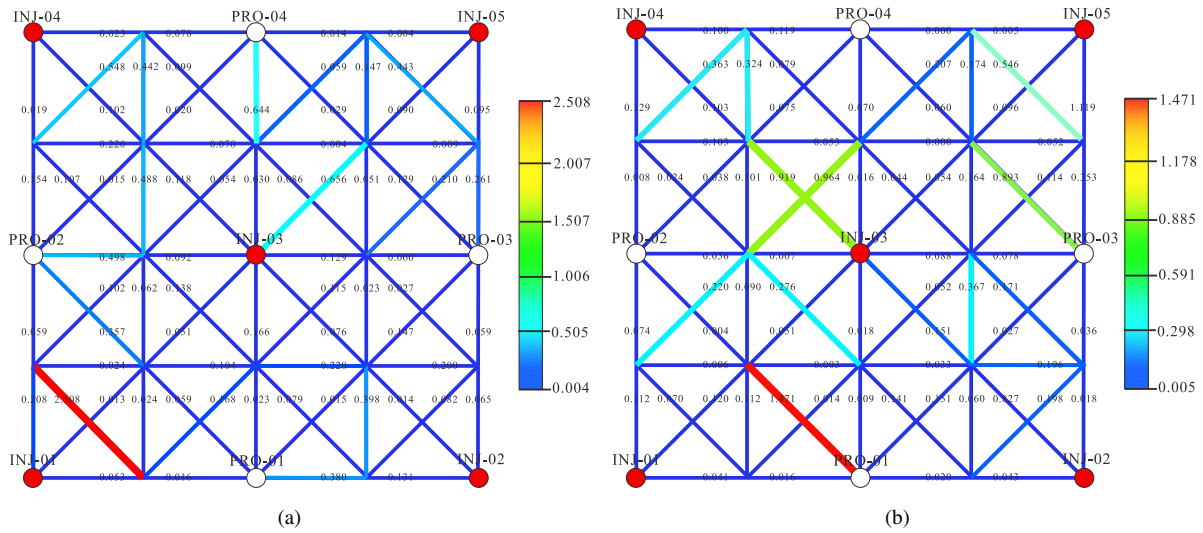


Fig. 14. Comparison of control PV of connection elements over (a) INSIM and (b) GPDF.

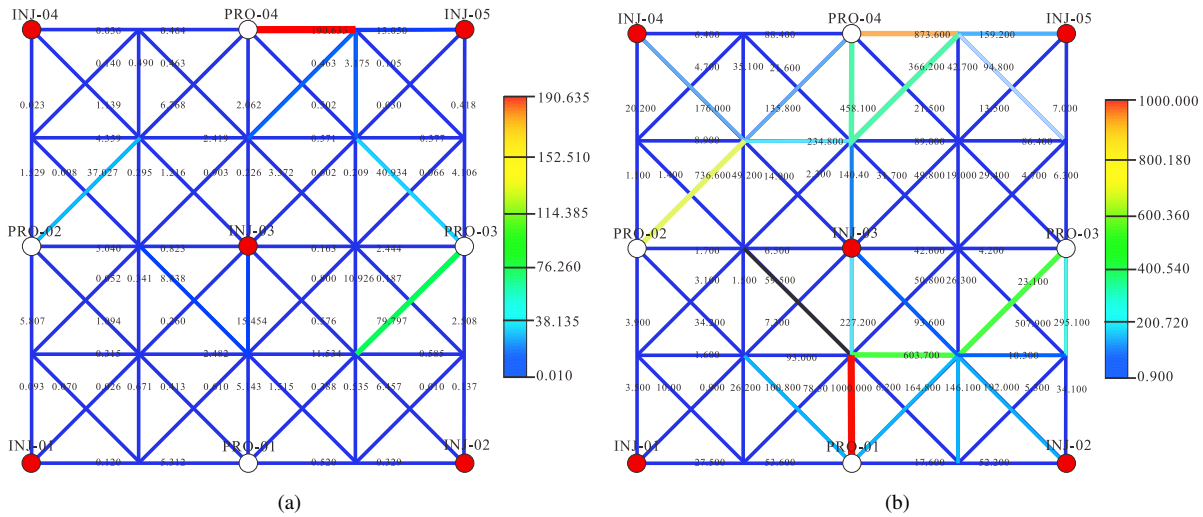


Fig. 15. Comparison of transmissibility of connection elements over (a) INSIM and (b) GPDF.

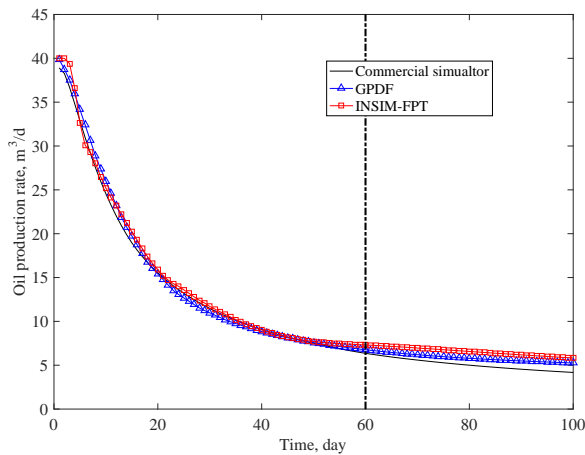


Fig. 16. History-matched single-well oil production rates over well PRO-01.

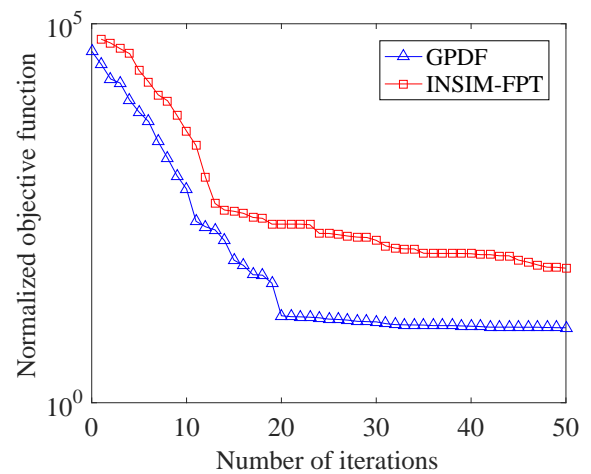


Fig. 17. Normalized objective function vs. number of iterations.

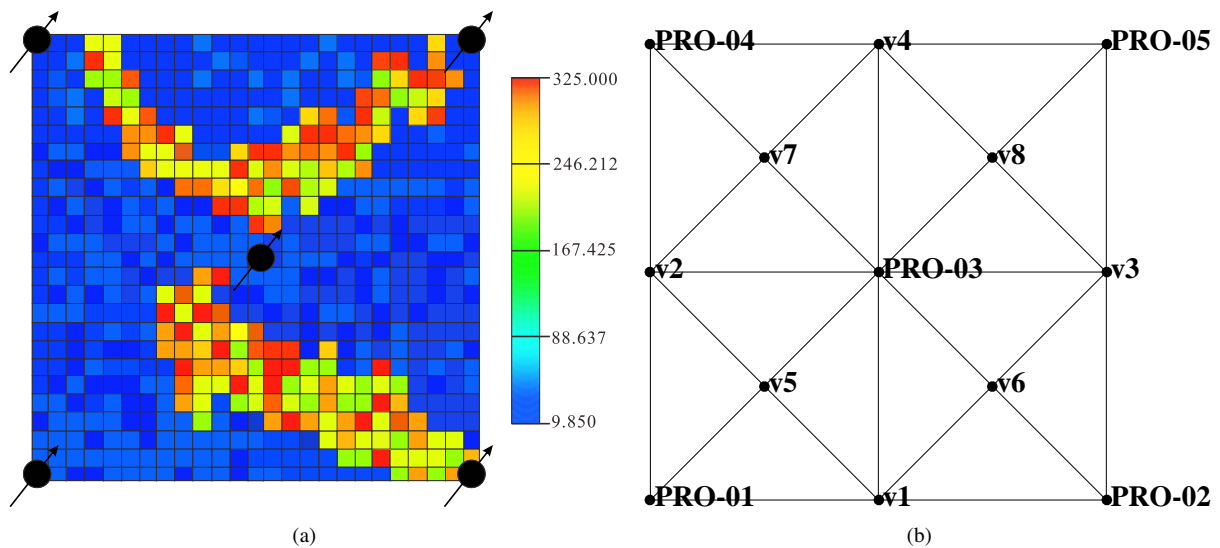


Fig. 18. Permeability distribution and network model of the reservoir model. (a) permeability profile of the grid-based reservoir model; (b) network model in GPDF.

Table 4. Physical properties used in example 4.

Properties	Values
Porosity	0.2
Oil volume factor	1.01
Grid size	25 m
Oil density	912 kg/m ³
Rock compressibility	1.07 × 10 ⁻⁴ MPa ⁻¹
Initial oil saturation	1
Oil compressibility	10 ⁻⁴ MPa ⁻¹
Initial reservoir pressure	25 MPa
Oil viscosity	12 cp
Thickness	20 m

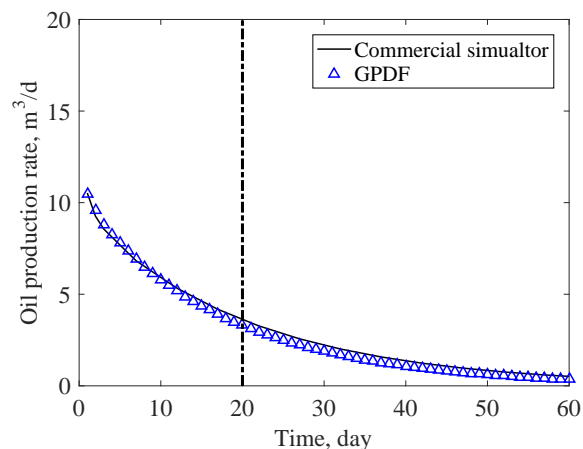


Fig. 19. Single-well oil production rates over Well #2.

ing, which is significantly less than the 625 (25 × 25) grid parameters in the ECLIPSE model in Fig. 19(a), which significantly reduces the multi solution of history matching and improves numerical simulation efficiency, which is a significant advantage of INSIM family methods.

4. Conclusions and future work

This paper presents a GPDF for numerical simulation and history matching of reservoirs that combines flow physics and well history data effectively and serves as a reference for constructing an effective physics-based data-driven model in other fields of science and engineering. Throughout the study, four major findings are drawn:

(1) This paper analyses the primary contents of the INSIM-family methods and analyzes their core concepts, highlighting three limitations of the INSIM family methods: poor generalization of flow models, inaccurate estimation of the upstream flux of 1D connection elements, and the unphysical constraint on the sum of the volumes of each connection element.

(2) Using two-phase oil-water flow as an example, GPDF solves FVM based fully implicit scheme of flow equations to yield all flow quantities concurrently, including pressure, saturation, component concentration, etc. It is explained that in GPDF, each 1D connection element retains two distinctive characteristic parameters, and the cross-sectional area and permeability of the grid on the same connection element are equal.

(3) The constraint on the sum of control volumes of each connection element should be unequal rather than equal in the data-driven (i.e., history matching) process of GPDF, and the total volume of all connection elements after history matching can be interpreted as the effective displacement area of reservoir development. This statement is illustrated numerically using a reservoir example with a high permeability zone.

(4) Numerous numerical examples are used to demonstrate

are only 56 (28 × 2) characteristic parameters in history match-

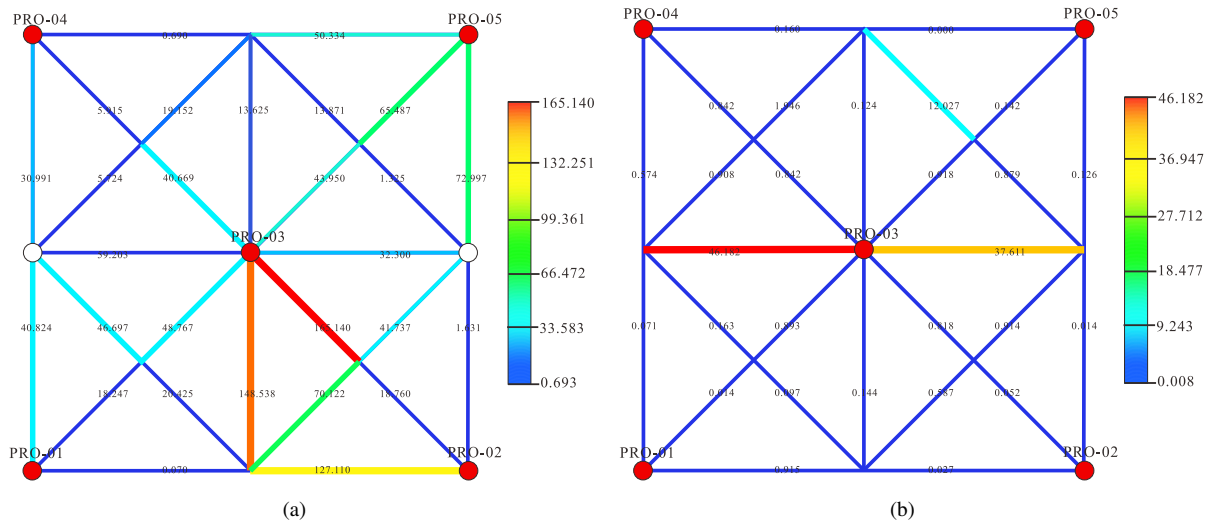


Fig. 20. Characteristic parameters of connection elements after history matching. (a) Transmissibility; (b) Control PV.

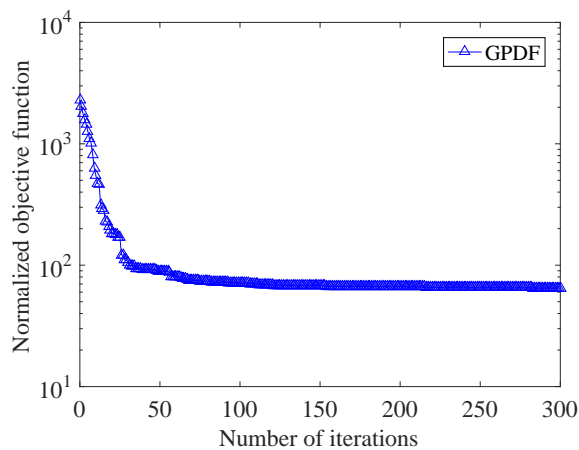


Fig. 21. Normalized objective function.

the superior performance of the presented GPDFs in terms of history matching and production forecast, as well as the good generalization of the GPDF flow models.

It is worth noting that, in comparison to the traditional numerical simulation method, this method retains a limitation, namely that the GPDF and INSIM family methods must still rely on the drive of history data from wells to obtain accurate parameter values for the connection elements. Thus, determining how to provide more accurate initial values for connection element parameters to increase the accuracy of the initial GPDF model's calculation, in order to act as a forward numerical simulator without history matching or to improve the performance of history matching with local convergence characteristics is an important future work. Additionally, we believe that there are no significant technical difficulties in expanding the GPDF from 2D to 3D, and that the dimension expansion is primarily determined by whether to consider vertical flow only at the well node when no connection element exists between two nodes in different vertical layers, or to consider gravity flow along connection elements when

connection elements exist between nodes in different layers, or to adopt a more complex approach. As a result, developing a 3D GPDF that is capable of high performance may be also a worthwhile future work.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (No. 52104017) and Cooperative Innovation Center of Unconventional Oil and Gas (Ministry of Education & Hubei Province), Yangtze University (No. UOG2020-17).

Conflict of interest

The authors declare no competing interest.

Open Access This article is distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC-ND) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

References

- Albertoni, A., Lake, L. W. Inferring interwell connectivity only from well-rate fluctuations in waterfloods. *SPE Reservoir Evaluation & Engineering*, 2003, 6(1): 6-16.
- Al-Yousef, A. A. Investigating statistical techniques to infer interwell connectivity from production and injection rate fluctuations. Austin, The University of Texas at Austin, 2006.
- Amirsardari, M., Dabir, B., Naderifar, A. Development of a flow based dynamic gridding approach for fluid flow modeling in heterogeneous reservoirs. *Journal of Natural Gas Science and Engineering*, 2016, 31: 715-729.
- Borregales, M., Møyner, O., Krogstad, S., et al. Data-driven models based on flow diagnostics. Paper ECMOR XVII 35122 Presented at European Association of Geoscientists & Engineers, Virtual, Online Event, 14-17 September, 2020.
- Cao, F., Luo, H., Lake, L. W. Oil-rate forecast by inferring fractional-flow models from field data with Koval method

- combined with the capacitance/resistance model. *SPE Reservoir Evaluation & Engineering*, 2015, 18(4): 534-553.
- Cardoso, M. A., Durlafsky, L. Use of reduced-order modeling procedures for production optimization. *SPE Journal*, 2010, 15(2): 426-435.
- Ding, D. Y., Wu, Y., Jeannin, L. Efficient simulation of hydraulic fractured wells in unconventional reservoirs. *Journal of Petroleum Science and Engineering*, 2014, 122: 631-642.
- Dinh, A. V., Tiab, D. Inferring interwell connectivity from well bottomhole-pressure fluctuations in waterfloods. Paper SPE 106881 Presented at the Production and Operations Symposium, Oklahoma City, Oklahoma, 31 March-3 April, 2007.
- Dinh, A. V., Tiab, D., Interpretation of interwell connectivity tests in a waterflood system. Paper SPE 116144 Presented at the SPE Annual Technical Conference and Exhibition, Denver, Colorado, 21-24 September, 2008.
- Griffiths, I. B. The extreme Spearman rank correlation coefficient in the characterization of the north buck draw field. Austin, University of Texas at Austin, 1998.
- Guo, Z., Reynolds, A. C. INSIM-FT in three-dimensions with gravity. *Journal of Computational Physics*, 2019, 380: 143-169.
- He, J., Durlafsky, L. J. Reduced-order modeling for compositional simulation by use of trajectory piecewise linearization. *SPE Journal*, 2014, 19(5): 858-872.
- Heffer, K. J., Fox, R. J., McGill, C. A., et al. Novel techniques show links between reservoir flow directionality, earth stress, fault structure and geomechanical changes in mature waterfloods. *SPE Journal*, 1997, 2(2): 91-98.
- Jansen, F., Kelkar, M. Non-stationary estimation of reservoir properties using production data. Paper SPE 38729 Presented at the SPE Annual Technical Conference and Exhibition, San Antonio, Texas, 5-8 October, 1997.
- Kiær, A., Lødøen, O., De Bruin, W., et al. Evaluation of a data-driven flow network model (flowNet) for reservoir prediction and optimization. Paper ECMOR XVII 35099 Presented at European Association of Geoscientists & Engineers, Virtual, Online Event, 14-17 September, 2020.
- Lake, L. W., Liang, X., Edgar, T. F., et al. Optimization of oil production based on a capacitance model of production and injection rates, Hydrocarbon economics and evaluation symposium. Paper SPE 107713 Presented at the Hydrocarbon Economics and Evaluation Symposium, Dallas, Texas, 1-3 April, 2007
- Lerlertpakdee, P., Jafarpour, B., Gildin, E. Efficient production optimization with flow-network models. *SPE Journal*, 2014, 19(6): 1083-1095.
- Liu, Y., Ding, Z., Ao, K., et al. Manufacturing method of large-scale fractured porous media for experimental reservoir simulation. *SPE Journal*, 2013, 18(6): 1081-1091.
- Martinez, A. N. A. Estimation of autocorrelation and its use in sweep efficiency calculation. Austin, The University of Texas at Austin, 1993.
- Maschio, C., de Carvalho, C. P. V., Schiozer, D. J. A new methodology to reduce uncertainties in reservoir simulation models using observed data and sampling techniques. *Journal of Petroleum Science and Engineering*, 2010, 72(1-2): 110-119.
- Naudomsup, N., Lake, L. W. Extension of capacitance-resistance model to tracer flow for determining reservoir properties. *SPE Reservoir Evaluation & Engineering*, 2019, 22(1): 266-281.
- Nguyen, A. P. Capacitance resistance modeling for primary recovery, waterflood and water-CO₂ flood. Austin, The University of Texas at Austin, 2012.
- Peng, X., Liang, B., Du, Z., et al. Practical simulation of multiporosity reservoirs through existing reservoir simulator. *Journal of Petroleum Science and Engineering*, 2017, 151: 409-420.
- Refunjol, B. T. Reservoir characterization of north buck draw field based on tracer response and production/injection analysis. Austin, The University of Texas at Austin, 1996.
- Ren, G., He, J., Wang, Z., et al. Implementation of physics-based data-driven models with a commercial simulator. Paper SPE 193855 Presented at the SPE Reservoir Simulation Conference, Galveston, Texas, 10-11 April, 2019.
- Sayarpour, M. Development and application of capacitance-resistive models to water/carbon dioxide floods. Austin, The University of Texas at Austin, 2008.
- Van Doren, J. F., Markovinović, R., Jansen, J. D. Reduced-order optimal control of water flooding using proper orthogonal decomposition. *Computational Geosciences*, 2006, 10(1): 137-158.
- Wood, D. A., Choubineh, A. Reliable predictions of oil formation volume factor based on transparent and auditable machine learning approaches. *Advances in Geo-Energy Research*, 2019, 3(3): 225-241.
- Wu, C., Wang, S., Yuan, J., et al. A prediction model of specific productivity index using least square support vector machine method. *Advances in Geo-Energy Research*, 2020, 4(4): 460-467.
- Yousef, A. A., Gentil, P. H., Jensen, J. L., et al. A capacitance model to infer interwell connectivity from production and injection rate fluctuations. *SPE Reservoir Evaluation & Engineering*, 2006, 9(6): 630-646.
- Zhao, H., Kang, Z., Zhang, X., et al., INSIM: A data-driven model for history matching and prediction for waterflooding monitoring and management with a field application. Paper SPE 173213 Presented at the SPE Reservoir Simulation Symposium, Houston, Texas, 23-25 February, 2015.
- Zhao, H., Xu, L., Guo, Z., et al. A new and fast waterflooding optimization workflow based on INSIM-derived injection efficiency with a field application. *Journal of Petroleum Science and Engineering*, 2019, 179: 1186-1200.
- Zhao, H., Xu, L., Guo, Z., et al. Flow-path tracking strategy in a data-driven interwell numerical simulation model for waterflooding history matching and performance prediction with infill wells. *SPE Journal*, 2020, 25(2): 1007-1025.