Petroleum Science 20 (2023) 3450–3460

Contents lists available at ScienceDirect

Petroleum Science

journal homepage: www.keaipublishing.com/en/journals/petroleum-science

Original Paper

Physics-informed neural network-based petroleum reservoir simulation with sparse data using domain decomposition

Jiang-Xia Han [a, b], Liang Xue [a, b, *], Yun-Sheng Wei [c], Ya-Dong Qi [c], Jun-Lei Wang [c], Yue-Tian Liu [a, b], Yu-Qi Zhang [d]

[a] National Key Laboratory of Petroleum Resources and Engineering, China University of Petroleum (Beijing), Beijing, 102249, China
[b] Department of Oil-Gas Field Development Engineering, College of Petroleum Engineering, China University of Petroleum, Beijing, 102249, China
[c] PetroChina Research Institute of Petroleum Exploration & Development, Beijing, 100083, China
[d] Lushang Oilfield Operation Area, Jidong Oilfield Company, PetroChina, Tangshan, 063000, Hebei, China

ARTICLE INFO

Article history:
Received 20 September 2022
Received in revised form
16 October 2023
Accepted 24 October 2023
Available online 27 October 2023

Edited by Yan-Hua Sun

Keywords:
Physical-informed neural networks
Fluid flow simulation
Sparse data
Domain decomposition

ABSTRACT

Recent advances in deep learning have expanded new possibilities for fluid flow simulation in petroleum reservoirs. However, the predominant approach in existing research is to train neural networks using high-fidelity numerical simulation data. This presents a significant challenge because the sole source of authentic wellbore production data for training is sparse. In response to this challenge, this work introduces a novel architecture called physics-informed neural network based on domain decomposition (PINN-DD), aiming to effectively utilize the sparse production data of wells for reservoir simulation with large-scale systems. To harness the capabilities of physics-informed neural networks (PINNs) in handling small-scale spatial-temporal domain while addressing the challenges of large-scale systems with sparse labeled data, the computational domain is divided into two distinct sub-domains: the well-containing and the well-free sub-domain. Moreover, the two sub-domains and the interface are rigorously constrained by the governing equations, data matching, and boundary conditions. The accuracy of the proposed method is evaluated on two problems, and its performance is compared against state-of-the-art PINNs through numerical analysis as a benchmark. The results demonstrate the superiority of PINN-DD in handling large-scale reservoir simulation with limited data and show its potential to outperform conventional PINNs in such scenarios.

1. Introduction

The modeling of flow and transport processes in subsurface formations is essential for various reservoir engineering applications, such as oil reservoir development and carbon sequestration. Quantifying flow mechanisms in these scenarios often requires using partial differential equations (PDEs) founded on laws of conservation. Traditionally, PDEs have been solved numerically using methods such as finite elements, finite differences, spectral, and meshless techniques. These methods have undergone significant development over the years, rendering them robust and flexible for solving complex subsurface flow and transport problems. However, numerical simulations can be computationally demanding for large-scale simulation problems in petroleum engineering, posing challenges during execution. Moreover, tasks like history-matching, sensitive analysis, and project design optimization require multiple simulation runs, leading to prohibitively inefficient computations for obtaining meaningful results (Ertekin and Sun, 2019). In recent years, data-driven methods have garnered significant attention and achieved remarkable progress in various domains such as natural language processing (Otter et al., 2021) and image classification (LeCun et al., 2015). Among these methods, artificial neural networks (ANNs) have emerged as one of the most important data-driven simulation techniques due to their universal approximation property and ability to accurately approximate any measurable function (Hornik et al., 1989). The versatility of ANNs extends to various science and engineering fields, demonstrating their potential in nonlinear universal approximation and data assimilation (Shanmuganathan, 2016; Abiodun et al., 2018).

* Corresponding author. National Key Laboratory of Petroleum Resources and Engineering, China University of Petroleum (Beijing), Beijing, 102249, China.
E-mail address: xueliang@cup.edu.cn (L. Xue).

https://doi.org/10.1016/j.petsci.2023.10.019

Notably, within the realm of petroleum engineering, ANNs have made substantial advancements in both forward modeling and other applications. Dong et al. (2019) proposed an enhanced artificial neural network model for predicting CO_2 minimum miscibility pressure based on the full composition of the crude oil and temperature. Erofeev et al. (2019) studied the applicability of various machine learning algorithms for predicting rock properties usually defined by geoscientists through specialized laboratory analysis. Moosavi et al. (2020) utilized multilayer perceptron neural networks to accurately identify reservoir models from pressure derivative curves derived from horizontal wells. Chung et al. (2020) proposed a pore-scale finite volume solver to predict permeability on digital cores by solving flow on micro-CT images. Kim et al. (2021) proposed an innovative data-integration method that uses an iterative-learning approach with a deep neural network coupled with a stacked autoencoder to address challenges encountered in many-objective history matching. Santos et al. (2021) tackled the limitation of modeling important geometries like fractures and vuggy domains accurately using a general multiscale deep learning model that can learn from porous media simulation data. Furthermore, Wang et al. (2021) outlined an integrated method that combines predictions of fluid flow with direct flow simulation, significantly reducing computation time without compromising accuracy. Alakeely and Horne (2022) examined the effectiveness of generative deep learning methods in predicting multiphase flow profiles of new wells in unseen locations using historical production data and a variational autoencoder algorithm. Dong et al. (2022) introduced a deep reinforcement learning based approach for automatic curve matching for well test interpretation, utilizing the double deep Q-network. Despite its remarkable advancements, the conventional data-driven approach inevitably faces several challenges. Firstly, the data-driven model is perceived as a "black box," since it lacks the incorporation of physical meaning of the dataset, leading to predictions that may be physically inconsistent or implausible (Karniadakis et al., 2021). Secondly, the robustness of the data-driven model may be poor, and its long-term prediction capabilities are weak. This can be attributed to the fact that generalizing and extrapolating beyond the parameter space of the included dataset is remain extremely challenging, as the models developed within artificial neural networks are bounded by the parameter space of the training dataset (Almajid and Abu-Al-Saud, 2022).

A recent and innovative neural network architecture that embodies this concept is known as 'physics-informed neural networks' (PINNs) (Raissi et al., 2019). By incorporating PDEs, boundary conditions, initial conditions, and other measurable prior knowledge into the loss function, PINNs seamlessly integrate data and physics laws to construct a physics-constrained loss function based on automatic differentiation. These algorithms not only improve the interpretability of machine learning models in terms of physical principles but also demonstrate robustness in handling flawed data, such as missing or noisy values, outliers, and other irregularities (Karniadakis et al., 2021). The performance of PINNs has been demonstrated in petroleum reservoir problems associated with subsurface flow and transport in porous media. Li et al. (2022) introduced a TGNN model that integrates high-fidelity numerical simulation data, fundamental physical laws, boundary conditions, initial conditions, expert knowledge, and engineering control terms in the loss function, enabling the solution of two-phase subsurface flow problems at the reservoir scale beyond the Buckley–Leverett equation. Daolun et al. (2021) proposed an improved physics-constrained PDE solution method by integrating potential features of the PDE into the loss functions to alleviate the strong nonlinear problem of the flow equation caused by the source-sink term in the single-phase homogeneous reservoir problem. Gasmi and Tchelepi (2021) utilized PINNs to tackle the two-phase immiscible flow problem governed by the Buckley–Leverett equation, achieving physical solutions by incorporating either a diffusion term into the partial differential equations or observed data. Almajid and Abu-Al-Saud (2022) implemented a physics-informed neural network technique that combines fluid flow physics and observed data to model the Buckley–Leverett problem. Cornelio et al. (2022) developed a neural network (NN) model to learn the physical model residual errors in simulation-based production prediction as a function of input parameters of an unconventional well. Hanna et al. (2022) developed a novel residual-based adaptive PINN and compared it with the residual-based adaptive refinement (RAR) method and a PINN with fixed collocation points. Wang et al. (2022) proposed a theory-guided convolutional neural network (TgCNN) framework that incorporates discretized governing equation residuals into the training of convolutional neural networks, to extend to two-phase porous media flow problems.

However, when applying PINNs to reservoir simulation with large-scale systems and production wells, existing research typically relies on training neural networks using high-fidelity numerical simulation data. The challenge arises from the scarcity of real reservoir data available for training purposes, which is often limited to information such as production rates and bottomhole flow pressures. Consequently, the accuracy of the physical information method may not always be guaranteed, when solving complex partial differential equations (PDEs) with large-scale systems and production wells, especially in the near-well zone where the pressure gradient is the largest.

In this study, we propose a physics-informed neural network based on domain decomposition (PINN-DD) to effectively utilize the sparse production data (bottom hole pressure and production data) of wells for reservoir simulation with a large-scale spatio-temporal domain. The introduced domain decomposition technique aims to harness the advantages of PINN in solving small-scale spatiotemporal domain problems to tackle the challenges posed by production wells and large-scale computational domain, which have very limited labeled data available. The computational domain is divided into two sub-domains: the well-containing and the well-free sub-domain, which are rigorously constrained by the governing equations, data matching, and boundary conditions. Specifically, the pressure at the wellbore radius determined by the Peaceman equation (Peaceman, 1978) serves as the inner boundary of the well-containing sub-domain.

2. Methodology

2.1. Multi-phase Darcy flows in petroleum reservoirs

In this section, we present a general oil–water two-phase Darcy flow model for porous media, which can be further simplified into a single-phase model. The model assumes that the fluids are slightly compressible and immiscible, and no mass transfer occurs between the phases. The governing equation is expressed as:

$$\begin{aligned} \nabla\left(\rho_{\mathrm{w}}\frac{kK_{\mathrm{rw}}}{\mu_{\mathrm{w}}}\nabla p_{\mathrm{w}}\right)+q_{\mathrm{w}}&=\frac{\partial(\phi\rho_{\mathrm{w}}S_{\mathrm{w}})}{\partial t}\\ \nabla\left(\rho_{\mathrm{o}}\frac{kK_{\mathrm{ro}}}{\mu_{\mathrm{o}}}\nabla p_{\mathrm{o}}\right)+q_{\mathrm{o}}&=\frac{\partial(\phi\rho_{\mathrm{o}}S_{\mathrm{o}})}{\partial t}\end{aligned} \tag{1}$$

where ρ_{w} and ρ_{o} are the density of water and oil, g/cm^3; k is the absolute permeability, mD; K_{rw} and K_{ro} are the relative permeability of oil and water, respectively; μ_{w} and μ_{o} are the water and oil viscosity, mPa s; p_{o} and p_{w} are the pressure of oil phase and water

phase, bar; S_w and S_o are the saturation of water and oil; q_w and q_o are the source/sink term of water and oil phase, m^3/d; ϕ is the porosity of porous media.

2.2. Physics-informed neural networks (PINNs)

In this subsection, we introduce the automatic differentiation (AD) algorithm and physics-informed neural network (PINN) model using fully-connected neural network (FC-NN) formulations. Automatic differentiation (AD), also known as computational differentiation, is a method for computing the derivatives of a function with respect to its input parameters automatically. AD works by representing the function as a graph of elementary operations, and then using the chain rule to propagate the derivatives through the graph (Lu et al., 2021). The core challenge addressed by AD is the computation of derivatives, gradients, and hessian matrices of complex functions. Nowadays, popular open-source machine learning libraries such as PyTorch and TensorFlow provide AD algorithms that are widely used in scientific computing and data analysis applications. By repeatedly applying the chain rule to simple arithmetic operations and elementary functions, AD can compute derivatives of any order with high precision.

We consider an initial-boundary problem in which the governing equation is expressible. The partial differential equations solved by neural networks can be presented in the following form:

$$f\left(t, x, y, u_t, \nabla u, \nabla^2 u, \ldots\right) = 0 \tag{2}$$

In the context of fluid flow problems in porous media, u typically denotes state variables such as the pressure or saturation within the reservoir. We suppose it can be represented by fully connected neural networks as u_θ:

$$u \approx u_\theta = W^{[n]} \times \sigma\left(W^{[n-1]} \times \sigma\left(\cdots\left(W^{[0]}[t,x,y]^{\mathrm{T}} + b^{[0]}\right)\cdots\right) + b^{[n-1]}\right) + b^{[n]} \tag{3}$$

where $W^{[n]}$ are weight matrices and $b^{[n]}$ are bias vectors for layer n; σ is a nonlinear activation function.

In PINNs, the loss function is comprised of several components, including the residual of the governing equation, initial and boundary conditions, and data matching. In the theoretical case, any dataset must satisfy the governing equation, meaning that the residual of the governing equation should approach zero. Consequently, the residual of the governing equation, as computed by Eq. (4), should be included as a regularization term in the neural network's loss function. The inclusion of the governing equation residual term ensures that the network satisfies the underlying physical principles, while the data matching and prior knowledge terms provide additional constraints to improve the accuracy of the predictions.

$$\mathrm{MSE}_{\mathrm{pde}} = \frac{1}{N_{\mathrm{p}}} \sum_{i=1}^{N_{\mathrm{p}}} \left[f\left(t_{\mathrm{p}}, x_{\mathrm{p}}, y_{\mathrm{p}}, u_{\theta ti}, \nabla u_{\theta i}, \nabla^2 u_{\theta i}, \ldots\right) \right]^2 \tag{4}$$

where $\{t_{\mathrm{p}}, x_{\mathrm{p}}, y_{\mathrm{p}}\}_{i=1}^{N_{\mathrm{p}}}$ denotes the collocation points of the residual of governing equation; N_{p} denotes the number of collection data.

Similarly, the mean square error associated with the initial condition, boundary condition, and data match can be expressed as:

$$\mathrm{MSE}_{\mathrm{BC}} = \frac{1}{N_{\mathrm{BC}}} \sum_{i=1}^{N_{\mathrm{BC}}} \left[u_{\theta i}(t_{\mathrm{BC}}, x_{\mathrm{BC}}, y_{\mathrm{BC}}) - u_i(t_{\mathrm{BC}}, x_{\mathrm{BC}}, y_{\mathrm{BC}})\right]^2 \tag{5}$$

$$\mathrm{MSE}_{\mathrm{IC}} = \frac{1}{N_{\mathrm{IC}}} \sum_{i=1}^{N_{\mathrm{IC}}} \left[u_{\theta i}(t_{\mathrm{IC}}, x_{\mathrm{IC}}, y_{\mathrm{IC}}) - u_i(t_{\mathrm{IC}}, x_{\mathrm{IC}}, y_{\mathrm{IC}})\right]^2 \tag{6}$$

$$\mathrm{MSE}_{\mathrm{data}} = \frac{1}{N_{\mathrm{data}}} \sum_{i=1}^{N_{\mathrm{data}}} \left[u_{\theta i}(t_{\mathrm{data}}, x_{\mathrm{data}}, y_{\mathrm{data}}) - u_i(t_{\mathrm{data}}, x_{\mathrm{data}}, y_{\mathrm{data}})\right]^2 \tag{7}$$

where $\{t_{\mathrm{BC}}, x_{\mathrm{BC}}, y_{\mathrm{BC}}\}_{i=1}^{N_{\mathrm{BC}}}$ and $\{t_{\mathrm{IC}}, x_{\mathrm{IC}}, y_{\mathrm{IC}}\}_{i=1}^{N_{\mathrm{IC}}}$ denote the collocation points of the boundary conditions and initial conditions, respectively; N_{IC} denotes the number of collection data; $u_i(t_{\mathrm{data}}, x_{\mathrm{data}}, y_{\mathrm{data}})$ and $u_{\theta i}(t_{\mathrm{data}}, x_{\mathrm{data}}, y_{\mathrm{data}})$ denote the label data and prediction data, respectively; N_{data} denotes the total number of label data.

Thus, the overall loss function can be formulated as follows:

$$\mathrm{Loss}(W, b) = \lambda_{\mathrm{pde}}\mathrm{MSE}_{\mathrm{pde}} + \lambda_{\mathrm{data}}\mathrm{MSE}_{\mathrm{data}} + \lambda_{\mathrm{IC}}\mathrm{MSE}_{\mathrm{IC}} + \lambda_{\mathrm{BC}}\mathrm{MSE}_{\mathrm{BC}} \tag{8}$$

where λ_{pde}, λ_{data}, λ_{IC}, λ_{BC} denote the weights of each term, and all take the value of 1 in this study. Currently, there is no systematic analysis in the literature of weight determination or optimization, and these weights are typically tuned by hand based on experience or trial and error. In this study, we introduce an approach to partially alleviate the cumbersome process of manual weight tuning, by ensuring that the value of each loss item falls within the same range. Specifically, we modify the training process by training the initial conditions and data matching in the first stage and subsequently adding the remaining items. This is feasible because the learning of the laws of physics and other components relies on the initial conditions.

2.3. PINNs based on domain decomposition

A typical reservoir problem comprises a set of equations that includes a continuity equation, initial conditions, outer boundary conditions, and inner boundary conditions. The inner boundary is represented by a wellbore, but the well radius is far smaller than the reservoir scale, and the region where pressure gradients are the largest is closest to a well, so it is often treated as a point source or a point sink in the calculation. However, the conventional PINN model cannot accurately solve the PDEs with source-sink terms (production and injection wells), unless the spatiotemporal fields are small or have a large amount of labeled data to guide the training process.

As illustrated in Fig. 1, the computational domain is divided into two distinct sub-domains: the small well-containing sub-domain 1 and the larger well-free sub-domain 2. The continuity assumption of the distribution of state variables at the partition interface connects the different domains. The spatiotemporal data of this interface are contained in both sub-domains and trained simultaneously, which results in the interface being constrained by the governing equations and boundary conditions of both sub-domains. Therefore, the continuity condition is implicitly satisfied without requiring additional constraints on the interface. The well is served as the inner boundary of the well-containing domain, and the formation pressure at the well-bore radius can be calculated using the Peaceman equation, as determined by Eq. (9). The structure of the PINN-DD model is illustrated in Fig. 2.

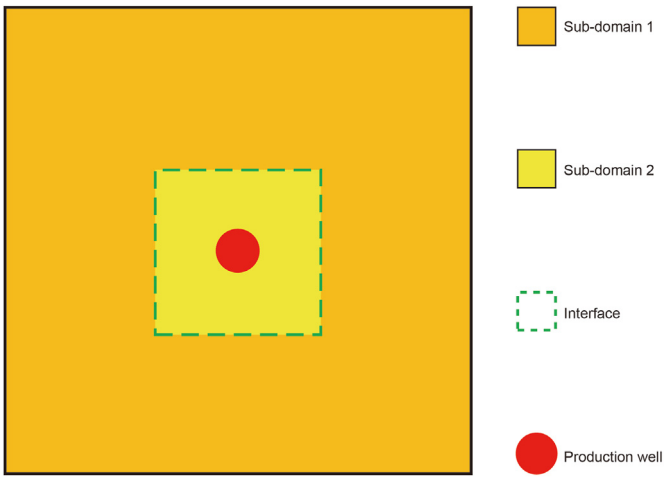

Fig. 1. The solution domain decomposition.

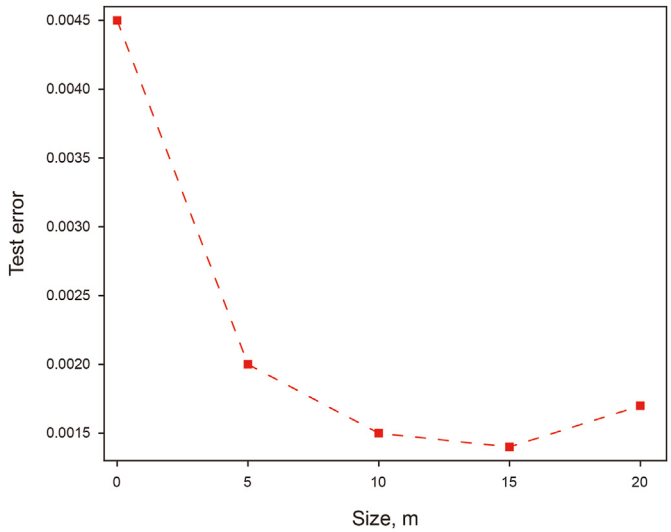

Fig. 3. Sensitivity analysis of interface size.

$$q = 2\pi\left(\frac{kh}{\mu}\right)\left(\frac{p_{\text{well}} - p_{\text{wf}}}{\ln\frac{r_e}{r_w} + S}\right) \tag{9}$$

where h is the thickness of the reservoir, m; μ is the fluid viscosity, mPa s; p_{well} is the well-containing grid pressure, bar; p_{wf} is the bottom hole flow pressure, bar; r_e is the equivalent radius, m; S is the skin effect, $S = 0$.

3. Results and discussion

In this section, we evaluate the performance of PINN-DD by assessing it on two distinct reservoir scenarios: a single-phase reservoir problem and an oil–water two-phase problem.

3.1. Sensitivity analysis of interface size

Accurately simulating pressure gradients near the wellbore requires carefully selecting the size of the interface. Large-size interfaces are advantageous for well-free domain calculations as it efficiently covers broader region, but cannot capture rapidly changing gradients close to the wellbore with desired accuracy. Conversely, smaller interfaces can simulate severe gradient changes and reduce the scale mismatch between the wellbore and well-containing domain, but at the cost of increased computational complexity in the well-free domain. Therefore, conducting a sensitivity analysis of interface size is essential to determine the optimal size that balances the computational efficiency of the well-free domain with the simulation precision of pressure gradients near the wellbore. To evaluate the impact of interface size on simulation accuracy, we conducted an analysis using a range of different interface sizes, considering the other parameter settings outlined in Section 3.2.

The results in Fig. 3 show that the test error decreases as the interface size increases until a certain point, beyond which it starts to increase. This observation aligns with the findings of the previous analysis. Based on considerations of computational efficiency and the sampling of collocation points data, we have chosen an interface size of 10 m for this study.

3.2. Single-phase reservoir problem

To assess the performance of PINN-DD, we conducted simulations for a single-phase problem with a production well. The computational domain was a 500-m square with no-flow boundaries. The production well was positioned at the center of the solution domain and operated at a constant rate of 50 m^3/d. The initial

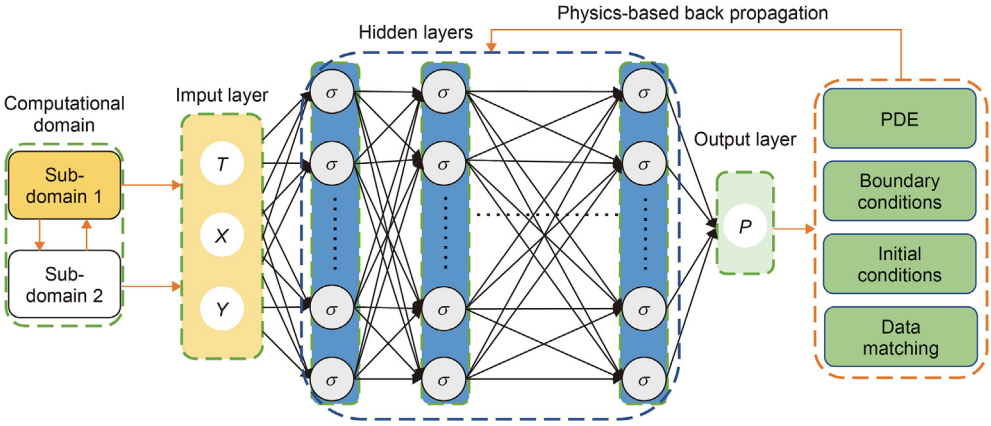

Fig. 2. The structure of the PINN-DD model.

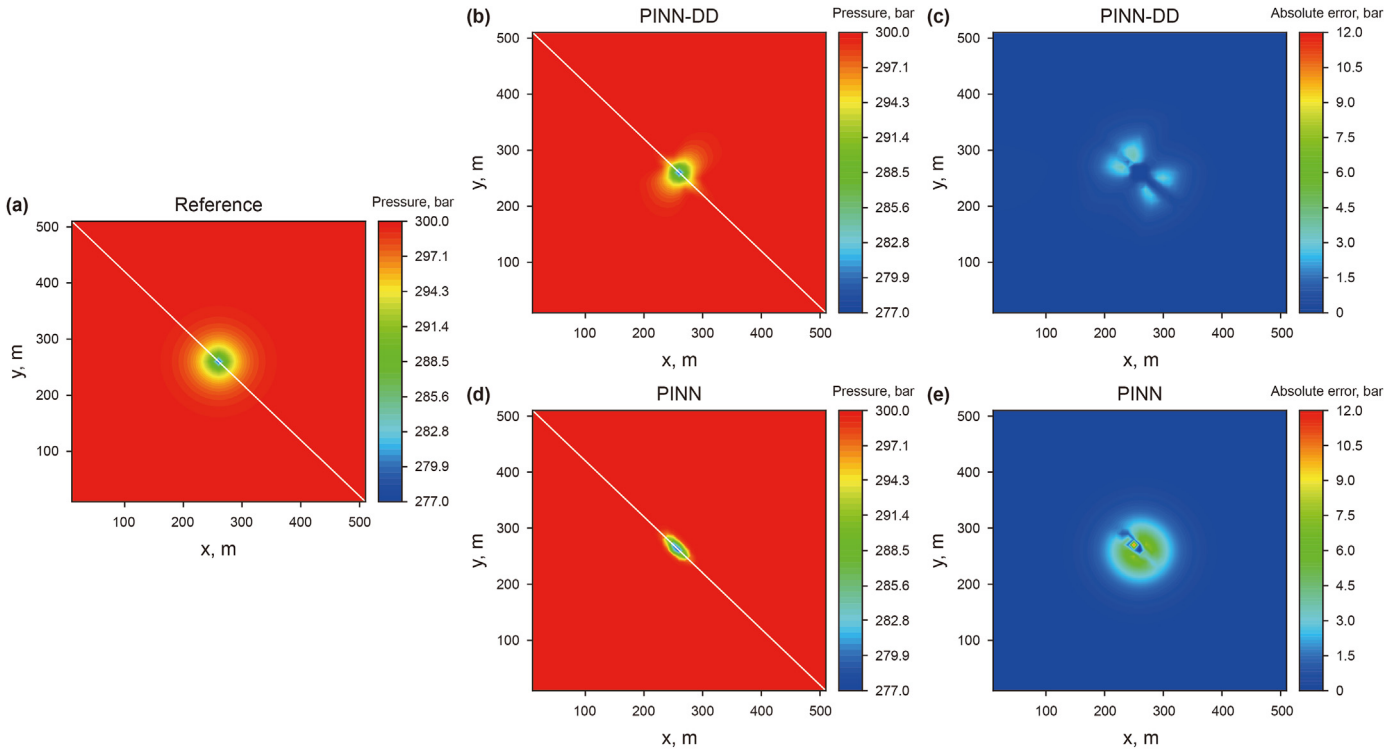

Fig. 4. Pressure fields of numerical simulation reference (**a**), PINN-DD (**b**, **c**), and PINN (**d**, **e**) at $t = 10$ d.

pressure was 300 bar, and the permeability, porosity, and C_f were 100 mD, 0.2, and 0.00045 bar^{-1}, respectively. A neural network with 15 hidden layers, each consisting of 100 neurons using the softplus activation function was employed. The Adam optimizer with a learning rate decay strategy was used to train the network. The coefficient of determination and the relative error L_2 were utilized to evaluate the accuracy, as follows:

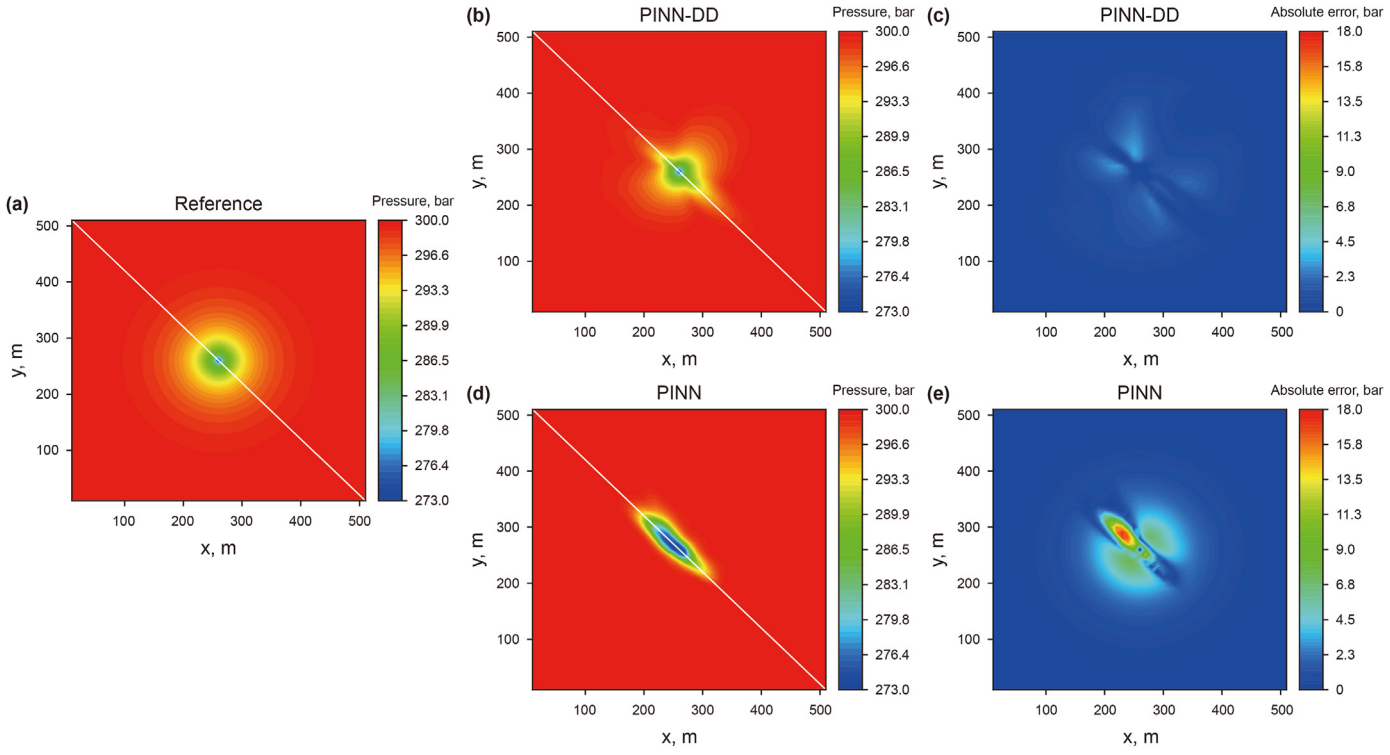

Fig. 5. Pressure fields of numerical simulation reference (**a**), PINN-DD (**b**, **c**), and PINN (**d**, **e**) at $t = 30$ d.

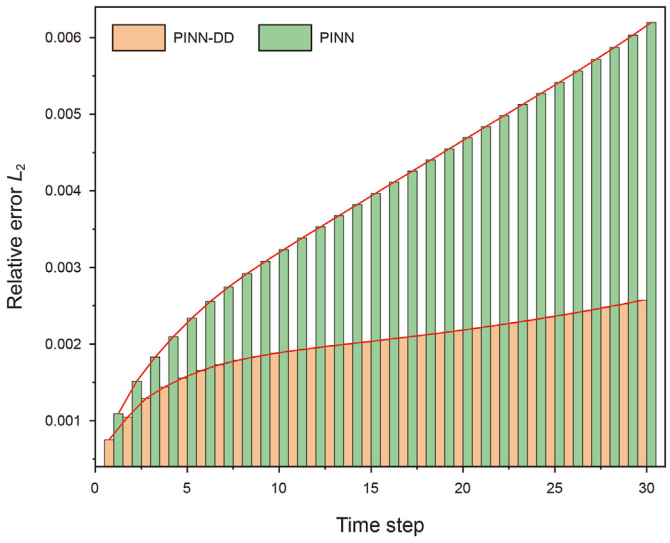

Fig. 6. Relative errors L_2 obtained by PINN-DD and PINN on the test dataset.

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (u_{\theta i} - u_i)^2}{\sum_{i=1}^{N} (u_i - \overline{u})^2} \tag{10}$$

$$L_2 = \frac{\|u_\theta - u\|_2}{\|u\|_2} \tag{11}$$

where $u_{\theta i}$ is the solution of PINNs; N is the number of evaluation points; u_i is the numerical solution of evaluation points; $\overline{u}$ is the average value of the numerical solution u.

The production rate and bottom-hole pressure data from a 30-day numerical simulation serve as the prior sparse label data. Subsequently, following the determination of network configurations, mathematical and geological models, as well as domain decomposition, the networks undergo training and testing. Notably, Figs. 4(b), 5(b), 4(d) and 5(d), illustrate the solved pressure fields of both PINN-DD and PINN at t = 10 d and t = 30 d, respectively. The absolute error of PINN-DD remains below 4 bar at t = 10 d and t = 30 d, in contrast to the PINNs, which amounts to 12 and 18 bar. As shown in Figs. 4 and 5, it becomes evident that the distribution pattern of the near-well zone of PINN-DD aligns more closely with the reference value, confirming that domain decomposition improves the simulation efficacy of the near-well zone.

Table 1
Summary of the evaluation metrics for the pressure on the diagonal obtained by PINN-DD and PINN.

Time, d	Model	R^2	MAPE, %
10	PINN-DD	0.994	0.0478
	PINN	0.687	0.292
30	PINN-DD	0.986	0.123
	PINN	0.734	0.467

The accuracy of results is further supported through the histogram of relative errors L_2 obtained by the PINN-DD and the PINN models on the test dataset, as illustrated in Fig. 6. Notably, while both approaches achieve comparable levels of relative errors L_2, the PINN-DD exhibits a heightened level of precision. Although PINN-DD also suffers from the error accumulation as the PINN model, we can see that the slope of the PINN-DD error over time is significantly smaller than that of PINN, proving a more stable generalization ability.

Fig. 7 and Table 1 show the pressure results obtained from PINN-DD and PINN along the diagonal line of the computational domain at t = 10 d and t = 30 d, providing a visual representation of pressure changes near the production well and offering insight into overall variations on a smaller-scale. The results obtained from PINN-DD perform better than those from PINN at both t = 10 d and t = 30 d, specifically at the near-well area with the largest pressure gradient. Domain decomposition, expertly applied in the PINN-DD approach, emphasizes the continuity of state variables near the well. Conversely, the PINN model endeavors to solve the entire computational domain as a whole unit, consequently, it struggles to precisely capture the intricate pressure fluctuations that arise within the near-well zone. Notably, inaccuracies in the pressure of the near-well zone can also propagate and impact the far-well zone, thereby exacerbating the limitations inherent in the traditional PINN methodology.

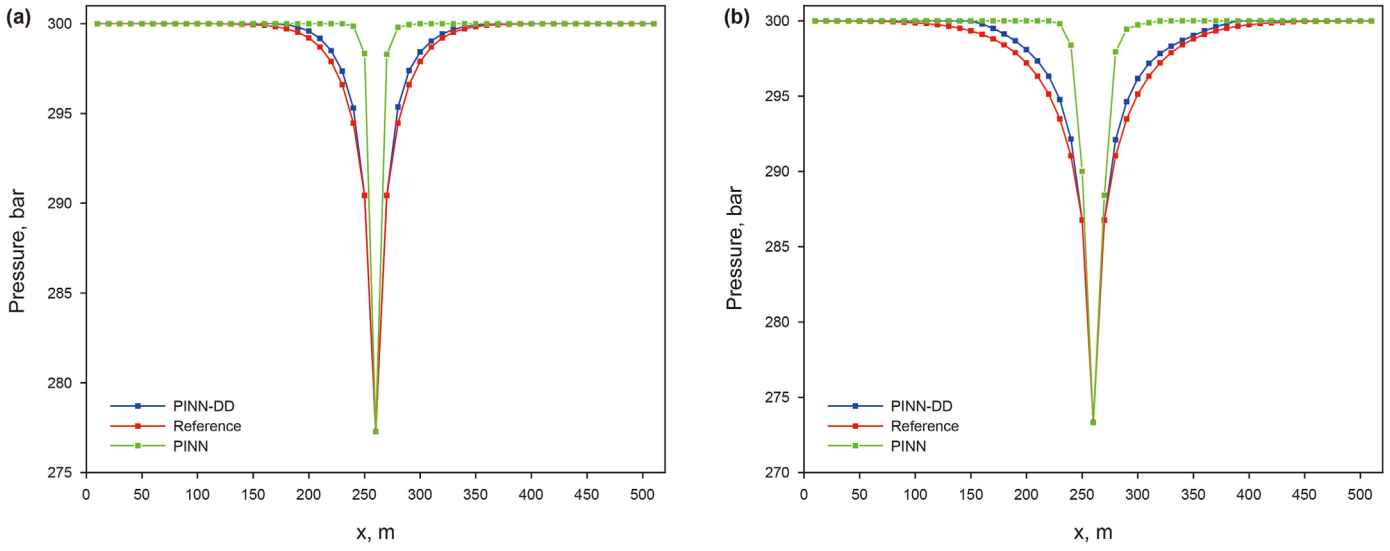

Fig. 7. The pressure obtained with PINN-DD and PINN on the diagonal line at t = 10 d (**a**) and t = 30 d (**b**).

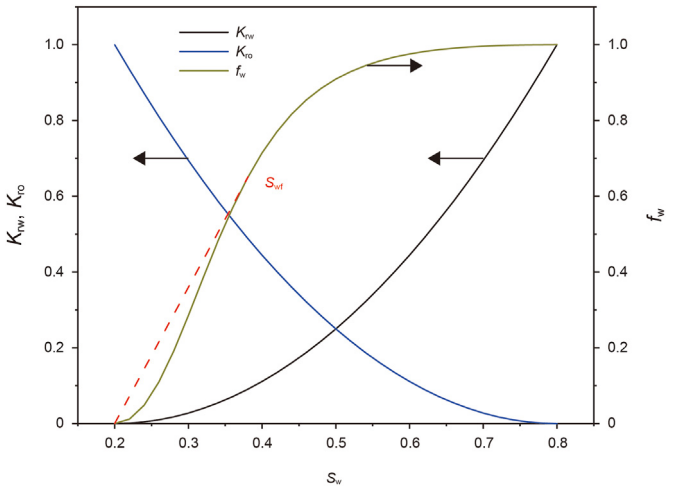

Fig. 8. The water, oil relative permeability, and water fractional flow function curves.

3.3. Oil–water two-phase problem

Unlike the single-phase problem, the governing equation for two-phase flow describes a coupled relationship between pressure and saturation. Consequently, this coupling presents significant challenges in solving the equation, highlighting its inherent nonlinearity.

In this investigation, we assumed that the fluid is both incompressible and immiscible, i.e., consistent with the 2D Buckley–Leverett two-phase problem. Based on these basic assumptions, we can construct a S_w-net to solve the saturation using PINNs based on Buckley–Leverett theory (Buckley and Leverett, 1942) and then solve the pressure in conjunction with the basic two-phase control equation using PINN-DD. The 2D Buckley–Leverett equation for the oil–water is written as:

$$\frac{\partial S_w}{\partial t} + \frac{Q}{\phi A(r)} \frac{\partial f_w}{\partial S_w} \frac{\partial S_w}{\partial r} = 0 \tag{12}$$

The initial and boundary conditions, respectively, are:

$$\text{Initial condication } S_w(t=0, r) = S_{wi} \tag{13}$$

$$\text{Boundary condition } S_w(t, r=r_w) = 1 - S_{or} \tag{14}$$

where S_w is the water saturation; f_w is the water fractional flow; r is the distance from the injection well; $A(r)$ is the cross-sectional area; ϕ is the porous medium porosity. The fractional flow function of water is defined as the ratio of water mobility to that of total mobility:

Table 2
Model parameters of the S_w-net.

Parameter	Value
Porosity	0.2
Permeability, mD	100
Irreducible water saturation	0.2
Residual oil saturation	0.2
Number of hidden layers	8
Number of neurons	50
Activation function	tanh

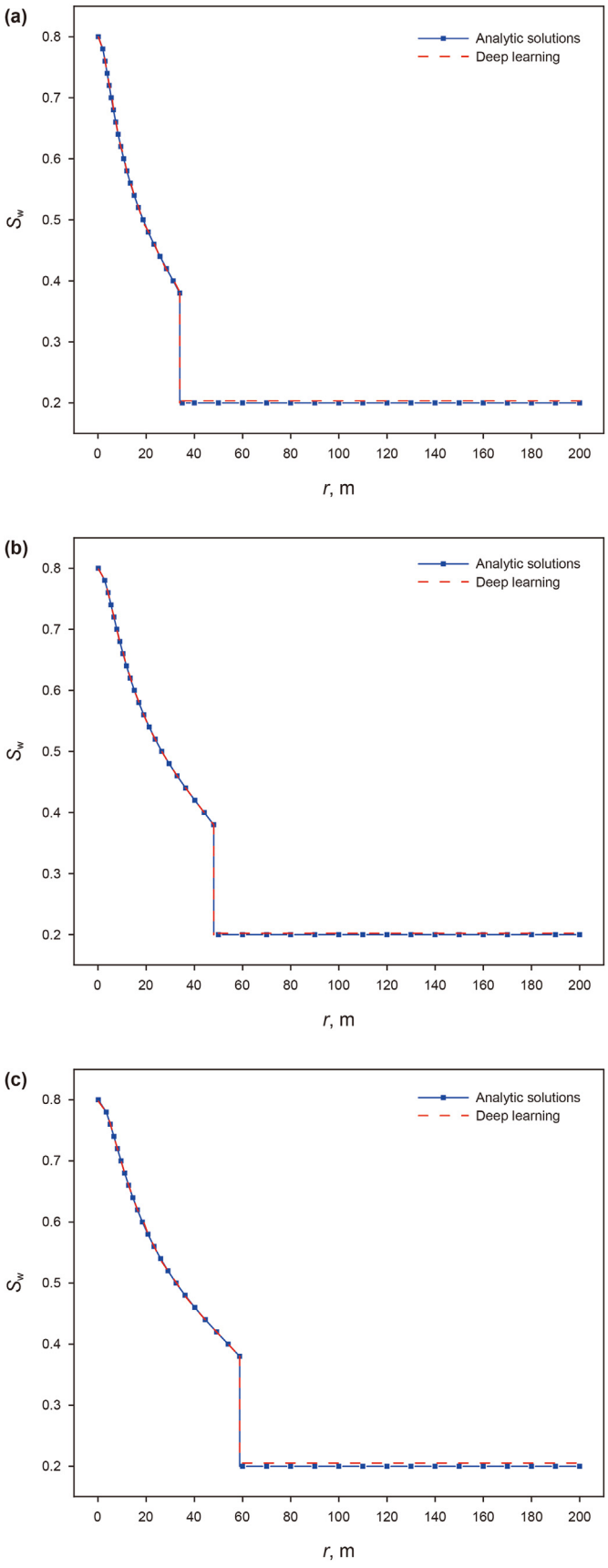

Fig. 9. The saturation fields of analytic solutions and S_w-net. (**a**) $t = 10$ d; (**b**) $t = 20$ d; (**c**) $t = 30$ d.

Table 3
The R^2 score of saturation prediction of S_w-net.

Time, d	R^2
10	0.9999
20	0.9999
30	0.9998

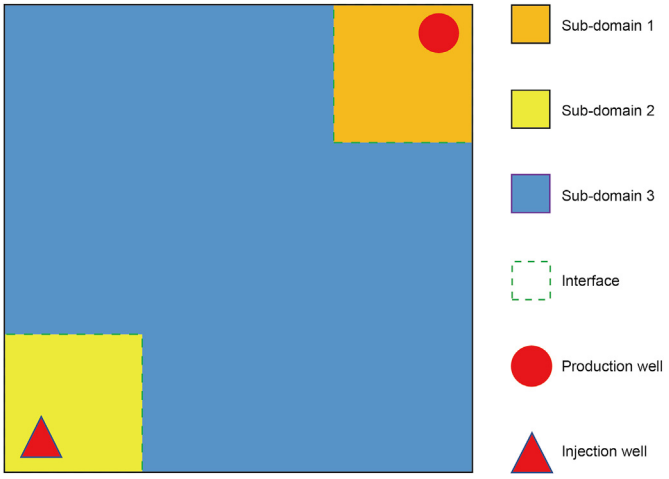

Fig. 10. Domain decomposition of the quarter five-spot problem.

$$f_w = \frac{\lambda_w}{\lambda_w + \lambda_o} = \frac{1}{1 + \frac{K_{ro}\mu_w}{K_{rw}\mu_o}} \tag{15}$$

Fig. 8 illustrates the relative permeability curves of water and oil. The saturation value of the oil–water front is a constant value that can be determined by drawing a tangent line to the water fractional flow curve, starting from the irreducible water saturation. The intersection point of the tangent line with the water fractional flow function curve represents the water saturation of the oil–water front, as denoted by the red dashed line in Fig. 8.

Numerous studies have explored the application of PINNs to solve the Buckley–Leverett problem (Fraces et al., 2020; Diab and Kobaisi, 2021; Rodriguez-Torrado et al., 2021; Xu et al., 2021; Diab et al., 2022). Given the wealth of literature on this topic, a comprehensive discussion of these works is beyond the scope of this study. In this study, we employ the S_w-net to solve the 2D Buckley–Leverett equation and use its pre-trained model as an auxiliary model in the process of solving the pressure field with PINN-DD. The pre-trained model provides the saturation values of each spatio-temporal point at each time step and further facilitates the determination of the coefficient values of the pressure term. First, the saturation of the oil–water front and its corresponding position are calculated using the saturation-displacement equation. The dataset is then used as data matching constraints, together with the saturation distribution in the single-phase region, to assist S_w-net in solving the Buckley–Leverett equation. Table 2 lists the model parameters of the NNs.

Fig. 9 shows the predicted saturation distributions of S_w-net at three time steps compared to the reference values. Based on Fig. 9 and the evaluation metrics in Table 3, it is evident that the S_w-net method accurately predicts the saturation distribution and captures the shock location.

The physical model to be solved is a 2D domain covering an area of 100 m × 100 m with four no-flow boundary conditions. The injection and production wells are located at coordinates (97.5, 97.5) and (2.5, 2.5), respectively. The computational domain can be divided into three sub-domains based on the presence of production wells, injection wells, or neither, as shown in Fig. 10. The initial pressure is 250 bar, while the initial water saturation is 0.2, and

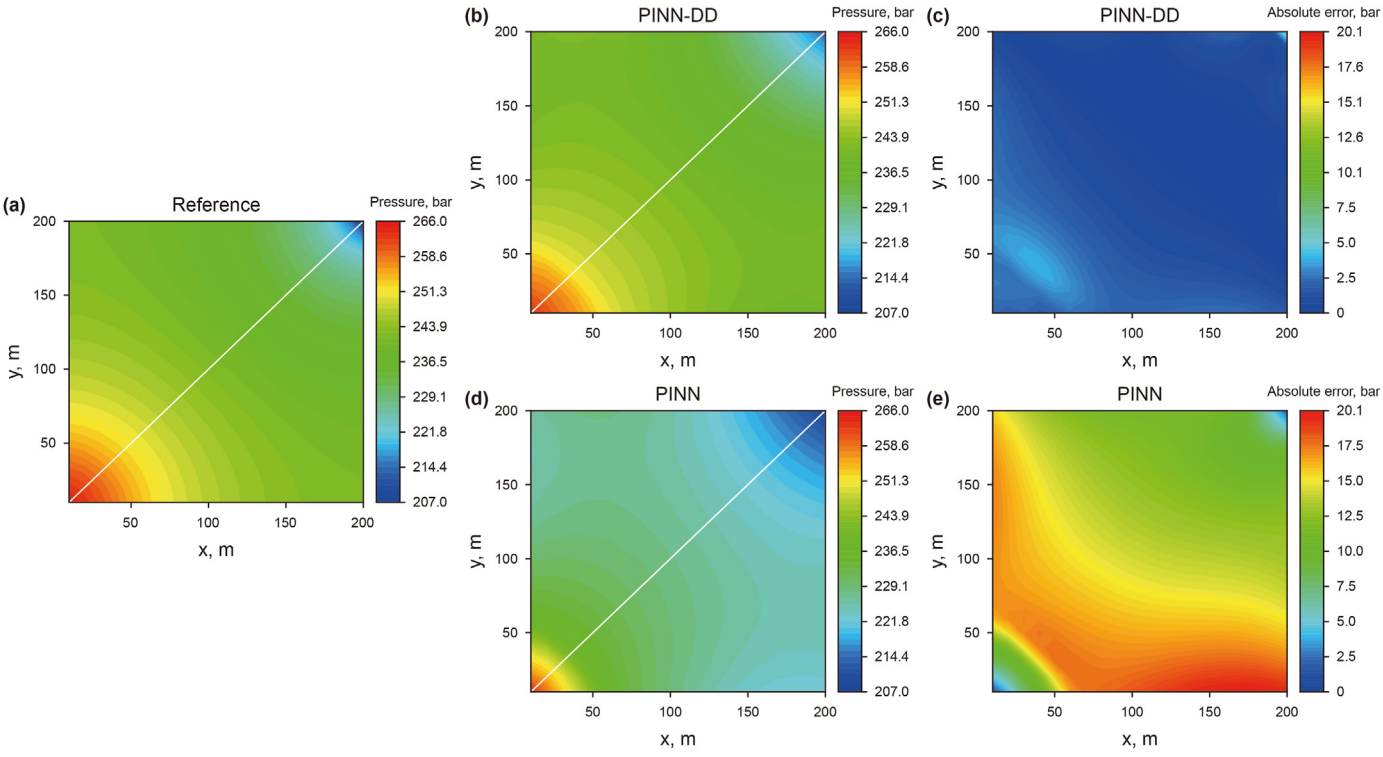

Fig. 11. Pressure fields of numerical simulation reference (**a**), PINN-DD (**b**, **c**), and PINN (**d**, **e**) at t = 10 d.

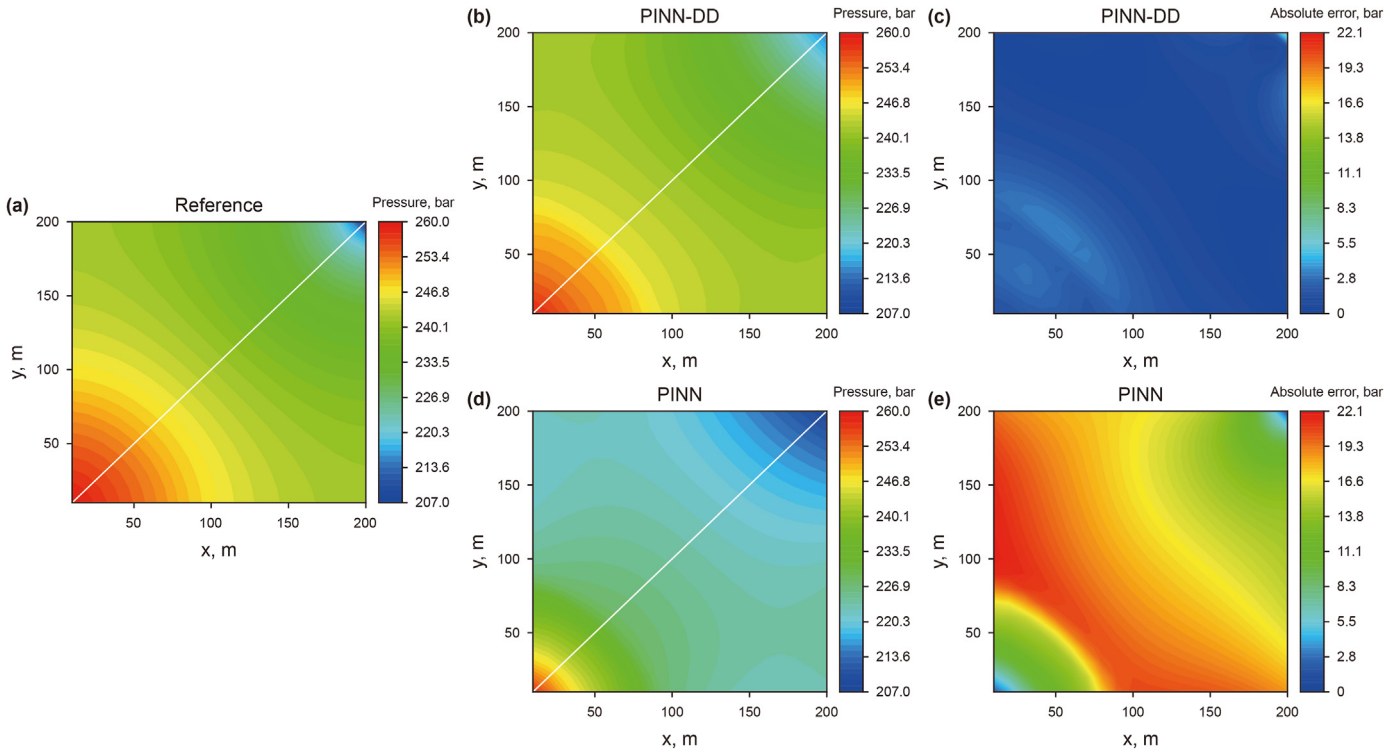

Fig. 12. Pressure fields of numerical simulation reference (**a**), PINN-DD (**b**, **c**), and PINN (**d**, **e**) at $t = 30$ d.

both irreducible water saturation and residual oil saturation are 0.2. The permeability is 100 mD, and the viscosity of water and oil is 0.3 and 3 mPa s, respectively. The injection and production wells operate at a constant liquid volume of 30 m^3/d and a constant well bottom pressure of 150 bar, respectively. The neural networks comprise 20 hidden layers, each with 50 neurons, and other network parameters are the same as in Subsection 3.2.

Figs. 11(b), 12(b), 11(d), and 12(d) show the pressure fields solved by both PINN-DD and PINN at $t = 10$ d and $t = 30$ d, respectively. The absolute error of PINN-DD remains below 8 bar, while the absolute error of PINN absolute error reaches 22 bar. The result shows that the PINN-DD has smaller errors, especially in the vicinity of the injection wells, indicating that domain decomposition can improve the accuracy of the large-scale computational domain problem with sparse production data.

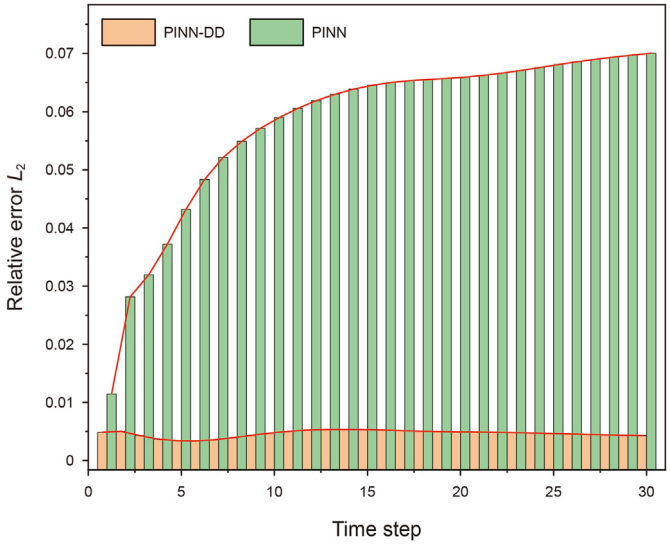

Fig. 13. Relative errors L_2 obtained by PINN-DD and PINN on the test dataset.

Fig. 13 compares relative errors L_2 between the PINN-DD and PINN models over 30 d. The results confirms that domain decomposition effectively mitigates the error accumulation. It is expected that the PINN-DD model may show an increasing error trend in later stages. Nevertheless, for the current 30-day simulation, the error accumulation effect in PINN-DD is not significant. Thus, domain decomposition improves the accuracy and stability of the PINN model in solving two-phase problems.

Fig. 14 and Table 4 show the pressure results obtained from both PINN-DD and PINN along the diagonal line at $t = 10$ d and $t = 30$ d. The results demonstrate that the PINN-DD curve closely agrees with the true value curve, while the standard PINN model exhibits significant discrepancies. The poor convergence of the PINN model can be attributed to the presence of discontinuous shock fronts in the saturation field and sharp changes in pressure gradients near the wellbore, which leads to inaccurate pressure field solutions near the injection well.

4. Conclusions

In this study, a physics-informed neural network based on domain decomposition (PINN-DD) is proposed to effectively utilize sparse production data from wells for reservoir simulation with large-scale systems. The novelty of this work is that, for the first time, the idea of the domain decomposition method is introduced into physics-informed-based large-scale reservoir simulation. It takes full account of the challenges posed by sparse data and the fact that the PINN model is good at solving small-scale spatio-temporal domain. By computational domain, it preserves the physical continuity of the near-well region where the pressure

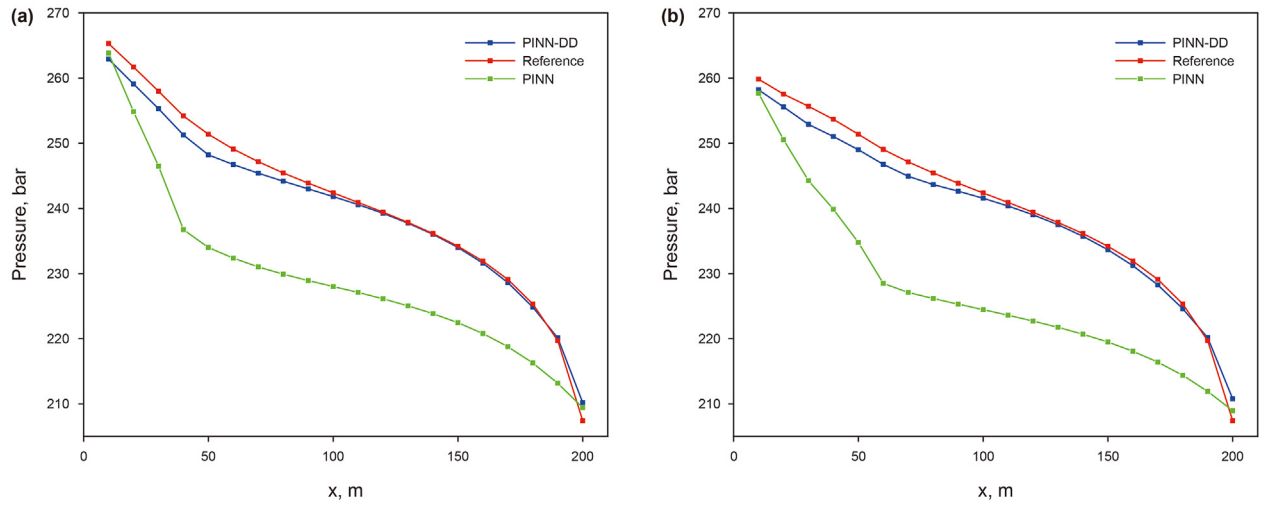

Fig. 14. The pressure with PINN-DD and PINN on the diagonal line at $t = 10$ d (**a**) and $t = 30$ d (**b**).

Table 4
Summary of the evaluation metrics for the pressure on the diagonal obtained by PINN-DD and PINN.

Time, d	Model	R^2	MAPE, %
10	PINN-DD	0.985	0.523
	PINN	0.1765	4.859
30	PINN-DD	0.9835	0.581
	PINN	<0	5.68

gradient is largest, and both the sub-domain and the interface are rigorously constrained by the governing equations, data matching, and boundary conditions.

The accuracy of the proposed method is evaluated on the single-phase and the oil–water two-phase problem with sparse production data, and its performance is compared against state-of-the-art PINNs through numerical analysis as a benchmark. In both cases, PINN-DD demonstrates its superiority of PINN-DD in handling large-scale reservoir simulation with sparse data, highlighting its potential to outperform conventional PINNs in such scenarios. In single-phase and two-phase scenarios at $t = 10$ d and $t = 30$ d, PINN-DD consistently achieves lower absolute pressure field errors compared to the standard PINN model. The domain decomposition technique helps reduce error accumulation, and although both PINN-DD and the standard PINN model are affected by error accumulation, PINN-DD experiences a gentler rate of error increase.

Several limitations of this study must be acknowledged. Firstly, while the PINN-DD method can capture the pressure dynamics system of large-scale reservoirs with limited data, it comes with the drawback of being computationally expensive and time-consuming. Secondly, even though the method only requires sparse production data, the data used in this study were obtained through numerical methods. Consequently, the absence of a history matching process, which corrects geological static parameters, represents another limitation of this study.

CRediT authorship contribution statement

Jiang-Xia Han: Conceptualization, Writing – original draft. **Liang Xue:** Conceptualization, Funding acquisition, Methodology. **Yun-Sheng Wei:** Methodology, Validation. **Ya-Dong Qi:** Data curation, Investigation. **Jun-Lei Wang:** Data curation, Formal analysis, Software. **Yue-Tian Liu:** Validation, Writing – review & editing. **Yu-Qi Zhang:** Data curation, Visualization.

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

This work was funded by the National Natural Science Foundation of China (Grant No. 52274048), Beijing Natural Science Foundation (Grant No. 3222037), the CNPC 14th Five-Year Perspective Fundamental Research Project (Grant No. 2021DJ2104) and the Science Foundation of China University of Petroleum - Beijing (No. 2462021YXZZ010).

References

Abiodun, O.I., Jantan, A., Omolara, A.E., Dada, K.V., Mohamed, N.A., Arshad, H., 2018. State-of-the-art in artificial neural network applications: a survey. Heliyon 4, e00938. https://doi.org/10.1016/j.heliyon.2018.e00938.

Alakeely, A., Horne, R., 2022. Simulating oil and water production in reservoirs with generative deep learning. SPE Reservoir Eval. Eng. 25 (4), 751–773. https://doi.org/10.2118/206126-PA.

Almajid, M.M., Abu-Al-Saud, M.O., 2022. Prediction of porous media fluid flow using physics informed neural networks. J. Petrol. Sci. Eng. 208, 109205. https://doi.org/10.1016/j.petrol.2021.109205.

Buckley, S.E., Leverett, M.C., 1942. Mechanism of fluid displacement in sands. Trans. AIME 146, 107–116. https://doi.org/10.2118/942107-G.

Chung, T., Da Wang, Y., Armstrong, R.T., Mostaghimi, P., 2020. CNN-PFVS: integrating neural network and finite volume models to accelerate flow simulation on pore space images. Transp Porous Med. 135, 25–37. https://doi.org/10.1007/s11242-020-01466-1.

Cornelio, J., Razak, S.M., Cho, Y., Liu, H., Vaidya, R., Jafarpour, B., 2022. Residual learning to integrate neural network and physics-based models for improved production prediction in unconventional reservoirs. SPE J. 27 (6), 3328–3350. https://doi.org/10.2118/210559-PA.

Daolun, L., Luhang, S., Wenshu, Z., Xuliang, L., Jieqing, T., 2021. Physics-constrained deep learning for solving seepage equation. J. Petrol. Sci. Eng. 206, 109046. https://doi.org/10.1016/j.petrol.2021.109046.

Diab, W., Kobaisi, M.A., 2021. PINNs for the Solution of the Hyperbolic Buckley-Leverett Problem with a Non-convex Flux Function arXiv preprint arXiv: 2112.14826.

Diab, W., Chaabi, O., Zhang, W., Arif, M., Alkobaisi, S., Al Kobaisi, M., 2022. Data-free and data-efficient physics-informed neural network approaches to solve the Buckley–Leverett problem. Energies 15, 7864. https://doi.org/10.3390/en15217864.

Dong, P., Chen, Z.-M., Liao, X.-W., Yu, W., 2022. A deep reinforcement learning (DRL) based approach for well-testing interpretation to evaluate reservoir parameters. Petrol. Sci. 19, 264–278. https://doi.org/10.1016/j.petsci.2021.09.046.

Dong, P., Liao, X., Chen, Z., Chu, H., 2019. An improved method for predicting CO_2 minimum miscibility pressure based on artificial neural network. Adv. Geo-

Energy Res. 3, 355–364. https://doi.org/10.26804/ager.2019.04.02.
Erofeev, A., Orlov, D., Ryzhov, A., Koroteev, D., 2019. Prediction of porosity and permeability alteration based on machine learning algorithms. Transp Porous Med. 128, 677–700. https://doi.org/10.1007/s11242-019-01265-3.
Ertekin, T., Sun, Q., 2019. Artificial intelligence applications in reservoir engineering: a status check. Energies 12, 2897. https://doi.org/10.3390/en12152897.
Fraces, C.G., Papaioannou, A., Tchelepi, H., 2020. Physics Informed Deep Learning for Transport in Porous Media. Buckley Leverett Problem. arXiv preprint arXiv: 2001.05172.
Gasmi, C.F., Tchelepi, H., 2021. Physics Informed Deep Learning for Flow and Transport in Porous Media arXiv:2104.02629 [physics].
Hanna, J.M., Aguado, J.V., Comas-Cardona, S., Askri, R., Borzacchiello, D., 2022. Residual-based adaptivity for two-phase flow simulation in porous media using physics-informed neural networks. Comput. Methods Appl. Mech. Eng. 396, 115100. https://doi.org/10.1016/j.cma.2022.115100.
Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Network. 2, 359–366. https://doi.org/10.1016/0893-6080(89)90020-8.
Karniadakis, G.E., Kevrekidis, I.G., Lu, L., Perdikaris, P., Wang, S., Yang, L., 2021. Physics-informed machine learning. Nat. Rev. Phys. 3, 422–440. https://doi.org/10.1038/s42254-021-00314-5.
Kim, J., Park, C., Ahn, S., Kang, B., Jung, H., Jang, I., 2021. Iterative learning-based many-objective history matching using deep neural network with stacked autoencoder. Petrol. Sci. 18, 1465–1482. https://doi.org/10.1016/j.petsci.2021.08.001.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539.
Li, J., Zhang, D., Wang, N., Chang, H., 2022. Deep learning of two-phase flow in porous media via theory-guided neural networks. SPE J. 27 (2), 1176–1194. https://doi.org/10.2118/208602-PA.
Lu, L., Meng, X., Mao, Z., Karniadakis, G.E., 2021. DeepXDE: a deep learning library for solving differential equations. SIAM Rev. 63, 208–228. https://doi.org/10.1137/19M1274067.
Moosavi, S.R., Vaferi, B., Wood, D.A., 2020. Auto-detection interpretation model for horizontal oil wells using pressure transient responses. Adv. Geo-Energy Res. 4, 305–316. https://doi.org/10.46690/ager.2020.03.08.
Otter, D.W., Medina, J.R., Kalita, J.K., 2021. A survey of the usages of deep learning for natural language processing. IEEE Transact. Neural Networks Learn. Syst. 32, 604–624. https://doi.org/10.1109/TNNLS.2020.2979670.
Peaceman, D.W., 1978. Interpretation of well-block pressures in numerical reservoir simulation (includes associated paper 6988). Soc. Petrol. Eng. J. 18, 183–194. https://doi.org/10.2118/6893-PA.
Raissi, M., Perdikaris, P., Karniadakis, G.E., 2019. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J. Comput. Phys. 378, 686–707. https://doi.org/10.1016/j.jcp.2018.10.045.
Rodriguez-Torrado, R., Ruiz, P., Cueto-Felgueroso, L., Green, M.C., Friesen, T., Matringe, S., Togelius, J., 2021. Physics-informed attention-based neural network for hyperbolic partial differential equations: application to the Buckley–Leverett problem. Sci. Rep. 12, 7557. https://doi.org/10.1038/s41598-022-11058-2.
Santos, J.E., Yin, Y., Jo, H., Pan, W., Kang, Q., Viswanathan, H.S., Prodanović, M., Pyrcz, M.J., Lubbers, N., 2021. Computationally efficient multiscale neural networks applied to fluid flow in complex 3D porous media. Transp Porous Med. 140, 241–272. https://doi.org/10.1007/s11242-021-01617-y.
Shanmuganathan, S., 2016. Artificial neural network modelling: an introduction. In: Shanmuganathan, S., Samarasinghe, S. (Eds.), Artificial Neural Network Modelling. Springer International Publishing, Cham, pp. 1–14.
Wang, N., Chang, H., Zhang, D., 2022. Surrogate and inverse modeling for two-phase flow in porous media via theory-guided convolutional neural network. J. Comput. Phys. 466, 111419. https://doi.org/10.1016/j.jcp.2022.111419.
Wang, Y.D., Chung, T., Armstrong, R.T., Mostaghimi, P., 2021. ML-LBM: predicting and accelerating steady state flow simulation in porous media with convolutional neural networks. Transp Porous Med. 138, 49–75. https://doi.org/10.1007/s11242-021-01590-6.
Xu, R., Zhang, D., Rong, M., Wang, N., 2021. Weak form theory-guided neural network (TgNN-wf) for deep learning of subsurface single- and two-phase flow. J. Comput. Phys. 436, 110318. https://doi.org/10.1016/j.jcp.2021.110318.