Computational Energy Science

Vol. 1, No. 1, p. 17-27, 2024

Original article

Deep learning surrogate model-based randomized maximum likelihood for large-scale reservoir automatic history matching

Wensheng Zhou[1,2], Wenhao Fu[3,4], Chen Liu[1,2], Kai Zhang[3,4,5]*, Jiahui Shen[3,4], Piyang Liu[5], Jinding Zhang[3,4], Liming Zhang[3,4], Xia Yan[3,4]

[1]*State Key Laboratory of Offshore Oil and Gas Exploitation, Beijing 100028, P. R. China*

[2]*CNOOC Research Institute Ltd., Beijing 100028, P. R. China*

[3]*State Key Laboratory of Deep Oil and Gas, China University of Petroleum (East China), Qingdao 266580, P. R. China*

[4]*School of Petroleum Engineering, China University of Petroleum (East China), Qingdao 266580, P. R. China*

[5]*School of Civil Engineering, Qingdao University of Technology, Qingdao 266520, P. R. China*

Keywords:
Automatic history matching
deep learning
surrogate model
randomized maximum likelihood

Cited as:
Zhou, W., Fu, W., Liu, C., Zhang, K., Shen, J., Liu, P., Zhang, J., Zhang, L., Yan, X. Deep learning surrogate model-based randomized maximum likelihood for large-scale reservoir automatic history matching. Computational Energy Science, 2024, 1(1): 17-27.
https://doi.org/10.46690/compes.2024.01.03

Abstract:
Automatic history matching in large-scale reservoir simulations poses significant challenges due to the complexity and uncertainty inherent in reservoir parameters. In this paper, we introduced a deep learning-based surrogate model, termed Convolution Recurrent Neural Network, for addressing these challenges. The Convolution Recurrent Neural Network leverages Convolution Neural Network and Recurrent Neural Network to extract spatial and temporal features respectively to approximate the intricate map between reservoir parameters and production data. And then, through the Randomized Maximum Likelihood method, the posterior distribution of reservoir parameters is sampled by optimizing a series of perturbed objective functions. The proposed framework several advantages, including its ability to handle high-dimensional data, capture complex reservoir dynamics, and efficiently calibrate uncertain parameters. Through comprehensive numerical experiments on both synthetic and real-world reservoir models, we demonstrate the efficacy of the approach in enhancing the efficiency and accuracy of automatic history matching in large-scale reservoir simulations.

1. Introduction

The accurate characterization of subsurface reservoirs is fundamental to optimizing hydrocarbon recovery in the oil and gas industry. However, achieving this accuracy is a daunting task due to the inherent complexity and uncertainty associated with reservoir parameters. History matching, the process of calibrating reservoir models to observed data, plays a pivotal role in oil and gas development.

Traditionally, history matching has been a labor-intensive and time-consuming process, often relying on manual adjustments to match simulated production data with observed field measurements. With the increasing size and complexity of reservoir models, there is a growing demand for automatic history matching techniques that can efficiently handle large-scale datasets while accounting for the inherent uncertainties in reservoir parameters.

History matching is a typical example of an ill-posed inverse problem, signifying that multiple parameter combinations can yield satisfactory matches to replicate past reservoir dynamics. To achieve a more comprehensive estimation of reservoir parameters, automatic history matching is often conceptualized as a sampling problem within a Bayesian framework. Nowadays, the pursuit of several optimal solutions or calibrating reservoir models to reliably assess the uncertainty of the predictions has emerged as the ultimate objective of

Yandy Scientific Press

*Corresponding author.
E-mail address: zhouwsh@cnooc.com.cn (W. Zhou); fwh230307@163.com (W. Fu); liuchen4@cnooc.com.cn (C. Liu); zhangkai@upc.edu.cn (K. Zhang); s23020055@s.upc.edu.cn (J. Shen); piyang.liu@qut.edu.cn (P. Liu).

Received January 8, 2024; revised February 15, 2024; accepted March 16, 2024; available online March 19, 2024.

history matching (Cancelliere et al., 2011). In light of this consideration, various history matching algorithms have been proposed over the past decades, including ensemble-based methods (Van Leeuwen and Evensen, 1996; Aanonsen et al., 2009; Emerick and Reynolds, 2013; Zhang et al., 2018) and Markov chain Monte Carlo methods (Liu and Oliver, 2003; Emerick and Reynolds, 2010; Li and Reynolds, 2020; Yan and Zhou, 2020). These Monte Carlo-based algorithms enable the acquisition of posterior distributions of uncertain parameters and provide assessments of uncertainty in posterior responses. However, these algorithms rely on iterative updates of reservoir parameters, necessitating extensive numerical simulations and thereby consuming substantial computational resources. Consequently, reducing the computational costs of history matching is imperative for reservoir production and management.

To further enhance the efficiency of history matching while maintaining accuracy, the integration of data-driven surrogate models has become a promising approach (Asher et al., 2015; Chen et al., 2023). Surrogate models, also known as proxy models, serve as computationally efficient approximations of complex reservoir simulators. These data-driven surrogate models encapsulate the underlying physics of reservoir behavior by probability approximation and can rapidly evaluate the response of the reservoir to different parameter configurations.

Recently, deep learning has experienced significant advancements and has been widely applied in various scientific and engineering domains, offering a novel approach to constructing alternative models for history matching. Traditional data-driven methods, such as kriging, polynomial regression, and k-nearest-neighbor (Hamdi et al., 2017; Wantawin et al., 2017; Yu et al., 2018) have prediction accuracy and computational efficiency limitations when facing high-dimensional and nonlinear data. To address this issue and promote the extensive utilization of surrogate models in history matching, deep learning has been incorporated to construct surrogate models. Treating the model parameters field as images, each grid point of the parameter field corresponds to a pixel in the image. Zhu and Zabaras (2018) used dense convolutional encoder-decoder networks to predict fluid velocity and pressure. Based on this, a training strategy combining regression loss and segmentation loss was proposed to better approximate the discontinuous saturation field (Mo et al., 2019). Tang et al. (2020) employed a residual U-Net network combined with convolutional long short-term memory recurrent networks to construct a surrogate model, capturing the spatiotemporal dynamics in high-dimensional nonlinear systems. Xiao et al. (2021) modified the residual U-Net network to improve model performance. Zhong et al. (2020) utilized a deep convolutional generative neural network (cDC-GAN) to establish a surrogate model, with time as a condition, model permeability field as input, and saturation distributions as output, moreover, production data obtained through Darcy's law and the principle of material balance. In image-to-image networks, production data cannot be directly obtained, which needs to be calculated based on the predicted pressure or saturation of the network.

In this study, we have combined Convolution Neural Network (CNN) and Recurrent Neural Network (RNN) to establish a surrogate model, achieving an end-to-end mapping from reservoir parameters to production data (Wantawin et al., 2017; Zhang et al., 2018; Ma et al., 2022a, 2022b). The neural network-based surrogate model in this paper is built using the open-access machine learning framework PyTorch (Paszke et al., 2019), and its training process can be viewed as an optimization procedure. By defining a loss function between model predictions and labels, the neural network parameters are gradually updated to improve prediction accuracy. Automatic history matching can also be regarded as an optimization problem. Therefore, once the neural network-based surrogate model is established, we can freeze the neural network parameters and directly compute the gradient of the automatic history matching objective function with respect to reservoir uncertain parameters using the Automatic Differentiation functionality within the PyTorch framework. We then utilize gradient descent to calibrate uncertain reservoir parameters. To effectively sample the posterior probability density function of uncertain reservoir parameters, we introduce the Randomized Maximum Likelihood (RML) (Kitanidis, 1986; Oliver et al., 1996) sampling framework within the proposed inversion framework, simultaneously optimizing a series of perturbed objective functions to estimate the posterior distribution of reservoir parameters.

The remaining sections of this paper are organized as follows. In section 2, we provide the fundamental theory behind the CNN and the RNN utilized in our data-driven surrogate model. Section 3 elaborates on the proposed Convolution Recurrent Neural Network surrogate model. Section 4 delineates the RML method, outlining the workflow of surrogate-based automatic history matching. Subsequently, Section 5 evaluates the proposed workflow on both two-dimensional synthetic reservoir models and three-dimensional large-scale reservoir models. Finally, we discuss and summarize the experimental results in section 6.

2. Related works

2.1 Deep residual convolution neural network

CNN play a crucial role in the field of deep learning and are particularly suited for processing data with spatial structure, such as images, speech, and text (LeCun et al., 2015). Generally, as the depth of the neural networks increase, the model can extract more features and tends to perform better. However, experiments have proved that this is not the case, when CNN reaches a certain depth, gradient vanishing and gradient explosion may occur if the number of network layers continues to increase, leading to setbacks in model training results. He et al. (2016) proposed a deep residual network, introducing residual blocks to overcome the problems caused by the increasing depth of the network.

The novelty of residual network is that uses shortcut connection structure, as shown in Fig. 1, x is the input of the neural network, $F(x)$ donates the residual mapping to be learned, and $H(x)$ is expected output, the residual learning unit takes the first two to get the output of the network, i.e., $H(x) = F(x) + x$. This type of skip connection enables an identity mapping between input and output features when $F(x)$ equals zero. The identity mapping directly passes the

input information to the output, allowing the entire network to learn only the difference between input and output, reducing parameter computation and preserving information integrity. Additionally, during backpropagation, the gradients of deep networks can be efficiently transported to shallower layers through the identity mapping, enabling the feedback of error information to earlier layers and effectively alleviating network degradation.

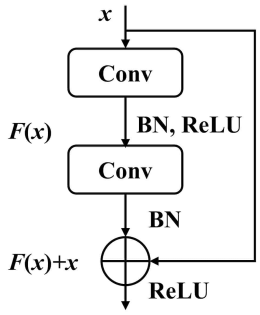

Fig. 1. The basic unit of a residual network.

2.2 Bidirectional long short-term memory unit

Long Short-Term Memory (LSTM) (Hochreiter and Schmidhuber 1997) is a variant of RNN designed to address the issue of gradient vanishing in traditional RNN. A basic LSTM unit structure is shown in Fig. 2, including the forget gate f_t, the input gate i_t, and the output gate o_t. x is the input at time t and h_{t-1} is the hidden state at time t-1. c_t is the cell state at time t, used to convey information.

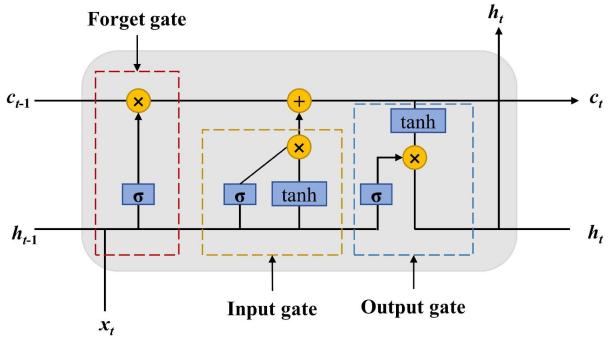

Fig. 2. The basic unit of LSTM.

The forget gate is responsible for determining which information to retain or discard, the input gate functions to update crucial information, and the output gate calculates the final output of the LSTM unit:

$$\begin{aligned} f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\ i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\ o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\ c_t &= f_t \cdot c_{t-1} + i_t \cdot \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \end{aligned} \tag{1}$$

where σ is the sigmoid activation function, W_f, W_i, W_o, W_c are the corresponding weights, b_f, b_i, b_o, b_c are the corresponding bias and $\tanh(\cdot)$ is the arctangent function, x_t is the input of the LSTM at time t.

The bi-directional LSTM (BiLSTM) (Graves et al., 2013) learns from both past and future data to overcome the limitation of LSTM that can only capture information in one direction, resulting in superior performance in time series prediction. The BiLSTM network is composed of two LSTM layers, one is forward, with inputs in the forward direction of the time series, and the other layer is backward, with inputs in the reverse time direction. The output of the BiLSTM network is determined jointly by these two layers, as shown in Fig. 3.

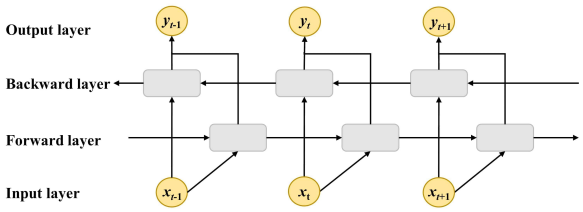

Fig. 3. The architecture of BiLSTM.

Where x and y are the input and output of the BiLSTM, in the forward layer, the first set of outputs are $\{z_1, z_2, \cdots, z_{k-1}\}$, while in the backward layer, the second set of outputs are $\{z_k, z_{k-1}, \cdots, z_1\}$, and these two sets of outputs are concatenated to obtain the final output.

3. Proposed surrogate model

As an effective technique, reservoir numerical simulation helps engineers understand the behavior of reservoir and predict reservoir production, enabling more accurate decision-making in reservoir development. In order to obtain reservoir models that approximate the actual reservoir, it usually requires hundreds of numerical simulations to update model parameters during the process of history matching. For complex and large-scale reservoirs, however, numerical simulation is time-consuming, leading to low computational efficiency. With the aim of solving this problem, surrogate model is constructed to replace the numerical simulation process.

3.1 General architecture

In numerical simulation, production data can be obtained by giving model parameters. For the purpose of expediting this process, surrogate model is constructed. The relationship between the input and output of the numerical simulation model can be described as:

$$F =: R^{H \times W \times N_x} \rightarrow R^{T \times N_d} \tag{2}$$

where $H \times W$ is the number of grids, N_x is the number of parameter fields, T is the timesteps, and N_d is the production data of different wells. Compared with numerical simulation, surrogate model has simple structures, ease for rapid predictions of reservoir production dynamics, and significant improvement in computational efficiency.

The proposed surrogate model utilizes convolutional and recurrent neural networks to establish a mapping relationship between model parameters and production data. CNN is used

for spatial feature extraction of model parameters and RNN is applied to handle time-varying production data. Details about these two parts will be provided below.

3.2 Network design for proposed model

As shown in Figs. 4 and 5, the model parameters are input into the network and initially processed through a convolutional layer and a max-pooling layer, then, they are passed through a series of operations involving residual blocks and max-pooling layers in an alternating fashion to extract spatial features, finally, the feature vector $\mathbf{z}$ is obtained through a fully-connected layer. In the RNN module, the spatial feature vectors $\mathbf{z}$ are fed into the first BiLSTM layer, which produces the hidden state h. The hidden state h then flows into the second BiLSTM layer to generate the final output y, representing the production data at different time steps.

3.3 Loss function and training procedure

An appropriate loss function can help models make accurate predictions. In the proposed surrogate model, we use Mean Absolute Error function as the loss function, which measures the average absolute difference between the predicted results and the actual values:

$$MAE = \frac{\sum_{i=1}^{n} |\mathbf{y}_i^{sim} - \mathbf{y}_i^{pre}|}{n} \tag{3}$$

where $\mathbf{y}_i^{sim}$ is the simulation data of ith sample, $\mathbf{y}_i^{pre}$ is the model prediction data of ith sample.

The entire training process is conducted in the Pytorch framework, and the training runs for 200 epochs. Adam algorithm (Kingma and Ba, 2014), as the optimizer, is utilized to update the model parameters based on the back-propagation method. The initial learning rate is set as 0.001 and then divided by 10 when the loss of the validation set stagnates.

4. Automatic history matching workflow with surrogate-based RML

4.1 Parameterization of uncertain parameters using principal component analysis

The significant uncertainty associated with large-scale reservoir parameters poses a crucial challenge for automatic history matching. With numerous geological and fluid properties involved, high-dimensional reservoir parameters constitute a complex multidimensional space. Within these high-dimensional parameters, redundant and correlated features often exist, presenting significant obstacles to the solution of automatic history matching. Consequently, compressing high-dimensional reservoir parameters into a lower-dimensional space for calibration becomes essential.

In the field of automatic history matching, commonly used parameter reduction techniques, also referred to as parameterization, include Principal Component Analysis (PCA) (Reynolds et al., 1996; Sarma et al., 2006), Discrete Wavelet Transform (Lu and Horne, 2000), and Discrete Cosine Transform (Jafarpour et al., 2010). Among these, PCA stands out as a prevalent dimensionality reduction method, demonstrating versatile advantages in handling high-dimensional data. PCA can effectively captures the primary trends of the data, by mapping it onto a new coordinate system where the variance of the data is maximized.

The first step in performing the PCA procedure is to generate a set of geological realizations. Then the N_r realizations need to be assembled into a centered data matrix:

$$\boldsymbol{Y} = [\boldsymbol{m}_1 - \bar{\boldsymbol{m}}\ \boldsymbol{m}_2 - \bar{\boldsymbol{m}} \cdots \boldsymbol{m}_{N_r} - \bar{\boldsymbol{m}}] \tag{4}$$

where $\boldsymbol{Y} \in \boldsymbol{R}^{N_m \times N_r}$, $\boldsymbol{m}_i$ denotes the ith realization, and $\bar{\boldsymbol{m}}$ is the mean of all N_r realizations. Then the singular-value decomposition is performed for $\boldsymbol{Y}/\sqrt{N_r - 1}$, which obtains:

$$\boldsymbol{Y} = \sqrt{N_r - 1}\boldsymbol{U}\Sigma\boldsymbol{V}^T = \sqrt{N_r - 1}\Phi\boldsymbol{V}^T = \Phi\boldsymbol{P} \tag{5}$$

where $\boldsymbol{U}$ and $\boldsymbol{V}$ are the $N_m \times N_m$ and $N_r \times N_r$ unitary matrices respectively, Σ is a $N_m \times N_r$ diagonal matrix, whose diagonal components are nonnegative singular values, Φ is a $N_m \times N_r$ basis matrix, and $\boldsymbol{P}$ is $N_r \times N_r$ a matrix whose column vectors are subject to standard normal distribution. Based on the energy criterion, the largest $N_l (N_l < N_r - 1)$ singular values are chosen to realize parameter dimension reduction and noise elimination. Then the new PCA realizations can be represented as:

$$\boldsymbol{m}(\xi) \approx \Phi_l \xi + \bar{\boldsymbol{m}} \tag{6}$$

where Φ_l is a $N_m \times N_l$ matrix, whose columns come from first the N_l columns of Φ, and $\xi \in R^{N_l \times 1}$ is the reduced-space variable subject to a standard normal distribution.

4.2 Randomized maximum likelihood method

The objective function of automatic history matching is usually constructed based on a Bayesian framework, which can be expressed as can be expressed:

$$\begin{aligned} O(\boldsymbol{m}) = \arg\min_{m} [&(G(\boldsymbol{m}) - \boldsymbol{d}_{obs})^T \boldsymbol{C}_D^{-1} (G(\boldsymbol{m}) - \boldsymbol{d}_{obs}) \\ &+ (\boldsymbol{m} - \boldsymbol{m}_p)^T \boldsymbol{C}_m^{-1} (\boldsymbol{m} - \boldsymbol{m}_p)] \end{aligned} \tag{7}$$

where $G(\cdot)$ represents the numerical simulation, $\boldsymbol{d}_{obs}$ represents the observation data, $\boldsymbol{C}_D$ is a diagonal matrix to measure the observed error, $\boldsymbol{m}_p$ and $\boldsymbol{C}_m$ are the mean and covariance of the prior model parameters.

The objective function consists of a prior probability density function and a data likelihood, i.e., model mismatch term and data mismatch term. The model mismatch term serves as a regularization mechanism to prevent history matching solutions from deviating significantly from geological knowledge. The data mismatch term is to minimize the difference between simulated data and observed data.

History matching is an ill-posed inverse problem characterized by the presence of multiple solutions; hence, optimizing a single objective function is insufficient. An ideal history matching solution should estimate the posterior distribution of reservoir parameters. To achieve this, we introduce the RML method, which attempts to sample the posterior distribution by optimizing a series of independent perturbed objective functions. The RML method can sample a reasonable posterior probability density function when the responses are nonlin-

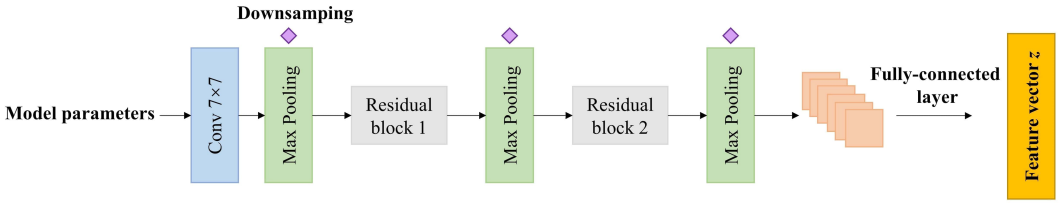

Fig. 4. Architecture of the residual convolutional network.

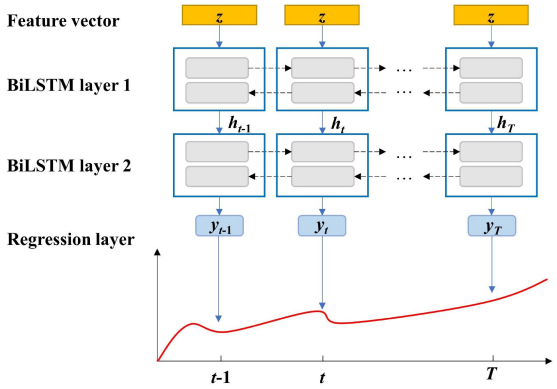

Fig. 5. Architecture of the bi-directional LSTM.

early related to the model parameters. Oliver et al. (2008) gives a concise RML process when hypothesizing that the model parameters $\boldsymbol{m}$ and observations noise are Gaussian. The process is as follows:

1) Sample $\boldsymbol{m}^*$ from the prior model Gaussian distribution.
2) Sample perturbed observations $\boldsymbol{d}^*_{obs}$ from the Gaussian distribution $N[\boldsymbol{d}_{obs}, \boldsymbol{C}_D]$.
3) Implement maximum likelihood estimation to obtain posterior variables $\boldsymbol{m}_{rml}$ by

$$\begin{aligned}\boldsymbol{m}_{rml} = \arg\min_{m}[&(G(\boldsymbol{m}) - \boldsymbol{d}^*_{obs})^T \boldsymbol{C}_D^{-1}(G(\boldsymbol{m}) - \boldsymbol{d}^*_{obs}) \\ &+ (\boldsymbol{m} - \boldsymbol{m}^*)^T \boldsymbol{C}_m^{-1}(\boldsymbol{m} - \boldsymbol{m}^*)]\end{aligned} \tag{8}$$

4) Repeat steps 1–3 to generate a set of posterior variables $\boldsymbol{m}_{rml}$.

Based on the PCA method, the relation between latent variable ξ and observations $\boldsymbol{d}_{obs}$ can be approximated as $\boldsymbol{d}_{obs} = G(\Phi_l \xi + \bar{\boldsymbol{m}}) + \boldsymbol{m} + \varepsilon$, where ξ follows a multivariate standard normal distribution, ε represents the observation noise. Therefore, the process of the RML method can be simplified as:

1) Sample ξ^* from the prior model Gaussian distribution $N[\boldsymbol{0}, \boldsymbol{I}]$.
2) Sample perturbed observations from the Gaussian distribution $N[\boldsymbol{d}_{obs}, \boldsymbol{C}_D]$.
3) Implement maximum likelihood estimation to obtain posterior variables ξ_{rml} by

$$\begin{aligned}\xi_{rml} = \arg\min_{\xi}[&(G(\boldsymbol{m}(\xi)) - \boldsymbol{d}^*_{obs})^T \boldsymbol{C}_D^{-1}(G(\boldsymbol{m}(\xi)) - \boldsymbol{d}^*_{obs}) \\ &+ (\xi - \xi^*)^T (\xi - \xi^*)]\end{aligned} \tag{9}$$

4) Repeat steps 1–3 to generate a set of posterior variables ξ_{rml}.

4.3 The automatic history matching workflow

Implementing reservoir simulation as an end-to-end surrogate model based on neural networks implies that the input reservoir model can be viewed as variables to be updated. In other words, leveraging neural networks to update the network parameters to fit the reservoir production data, then fixing these parameters to calibrate the input variables to match the observed data. Integrating the PCA and RML, the loss function of the automatic history matching is as follows:

$$\begin{aligned}O(\xi) = \frac{1}{2}[&(F(\boldsymbol{m}(\xi)) - \boldsymbol{d}^*_{obs})^T \boldsymbol{C}_D^{-1}(F(\boldsymbol{m}(\xi)) - \boldsymbol{d}^*_{obs}) \\ &+ (\xi - \xi^*)^T (\xi - \xi^*)]\end{aligned} \tag{10}$$

where $F(\cdot)$ represents the forward propagation of the neural network, and $\boldsymbol{d}^*_{obs}$ denotes the perturbed observations.

Above process entails transforming the complex relationships within reservoir systems into trainable parameters within the neural network architecture. These parameters are iteratively adjusted during the training process to minimize the mismatch between surrogate model predicted and observed data, ultimately capturing the dynamic behavior of the reservoir.

Through gradient backpropagation algorithms, we can effectively calibrate a set of uncertain parameters. The calibration of uncertain parameters is performed using the Adam algorithm integrated with damping strategy. Due to the fact that optimization based on surrogate model does not require running computationally time-consuming numerical simulations, optimization does not limit the maximum number of iterations and terminates when the relative rate of change of

the data loss term is less than 0.01.

5. Case study

In this section, we validated the effectiveness of the proposed automatic history matching framework on two cases of waterflooding reservoirs. For each instance, we first evaluate the predictive performance of the surrogate model on production data, and then calibrate the uncertain parameters of the reservoir based on the surrogate model.

5.1 Case 1: Heterogeneous waterflooding 2D reservoir model

The first case is a synthetic reservoir model with 60×60×1 grid blocks, and each block represents 20m×20m×4m. The porosity value is set to a constant of 0.2. We generate 2,000 prior permeability fields using Stanford Geostatistical Modeling Software (Remy et al., 2009), then perform simulations on these models to obtain corresponding production data. Among them, 1,400 models are used as the training set, 200 models are used as the validation set, and 400 models are used as the testing set. A model with the same geological description as the prior model but not included in the prior models is generated as the reference model. The reference log-permeability field is shown in Fig. 6. There are 9 production wells and 4 water injection wells, and all wells are controlled by bottom-hole pressure. In this case, the whole production period is 1,500 days and the interval for each time step is 30 days. The prediction objective of the surrogate model is the well oil production rate (WOPR) and the well water production rate (WWPR) of the 1,500 days production period.

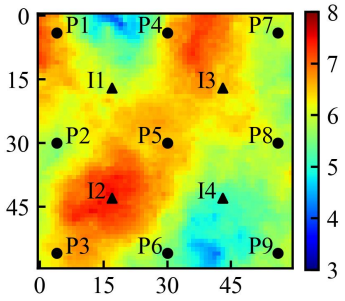

Fig. 6. Reference reservoir model: Case 1.

5.1.1 Prediction accuracy assessment

In this section, we mainly assess the performance of the proposed Convolution Recurrent Neural Network surrogate model for production data prediction. We use six different sample sizes (N_{train}=400, 600, 800, 1,000, 1,200, 1,400) to train the model and explore the impact of training sample size on prediction accuracy. The training epochs and training batches are set as 100 and 16, respectively. As mentioned earlier, the remaining 400 prior models are used as the testing set to evaluate the performance of the surrogate model. We select two commonly used performance metrics, the coefficient of determination (R^2) and the root-mean-square error (RMSE), for predictive performance evaluation. The formulas of the metrics are shown in Eqs. (11)-(12):

$$R^2 = 1 - \frac{\sum_{i=1}^{N_{test}} \|\boldsymbol{y}^i - \hat{\boldsymbol{y}}^i\|_2^2}{\sum_{i=1}^{N_{test}} \|\boldsymbol{y}^i - \bar{\boldsymbol{y}}^i\|_2^2} \tag{11}$$

and

$$RMSE = \sqrt{\frac{1}{N_{test}} \sum_{i=1}^{N_{test}} \|\boldsymbol{y}^i - \hat{\boldsymbol{y}}^i\|_2^2} \tag{12}$$

Respectively, where $\boldsymbol{y}^i$ denote the numerical simulation results, $\hat{\boldsymbol{y}}^i$ denote the surrogate model predictions, $\bar{\boldsymbol{y}}$ is the mean of the simulation results. The closer R^2 is to 1 and the lower RMSE value means the better surrogate quality.

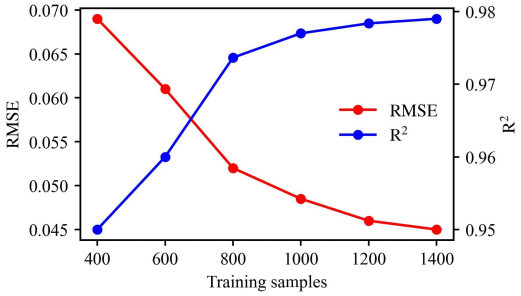

Fig. 7. Comparison of the evaluation results of the surrogate models on the test samples with different numbers of training samples.

Fig. 7 displays the evaluation results of RMSE and R^2 for test samples with different six training samples. The results indicate that with an increase in the number of training samples, the predictive accuracy of the surrogate model also improves. However, beyond a certain threshold of training samples, the marginal improvement in predictive accuracy diminishes significantly. At this point, further increasing the number of training samples yields diminishing returns, as the configuration of training samples itself requires extensive computational time for numerical simulations. Judging from both the RMSE and R^2 evaluation metrics, the inflection point for the predictive accuracy of the surrogate model is approximately between 800 and 1,200 training samples. Based on this observation, it can be roughly inferred that for the surrogate model proposed in this paper, the optimal number of training samples that balances predictive accuracy and computational costs lies between 800 and 1,200. In this particular case, we employ the surrogate model obtained from 1,000 training samples for automated history matching.

In order to provide a more intuitive demonstration of the predictive performance of the surrogate model, we concurrently present numerical simulation results of a randomly selected test case alongside the surrogate model predictions for evaluating the predictive quality of the surrogate model, as shown in Fig. 8. It can be observed that the surrogate model predictions for WOPR and WWPR closely match the numerical simulation results. These results indicate that the surrogate model proposed in this study can capture the intricate mapping relationship between heterogeneous reservoirs and production data.

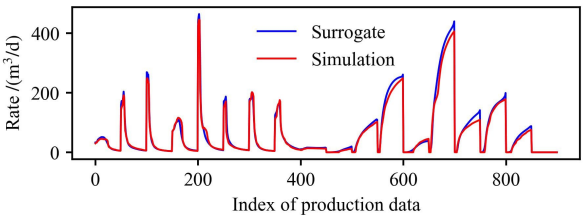

Fig. 8. Comparison of well rates from the numerical simulator and surrogate model.

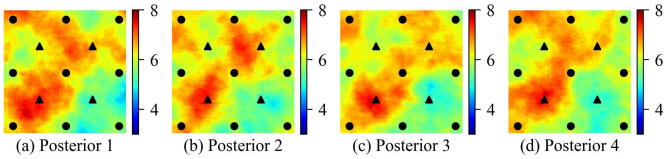

Fig. 9. Four posterior log-permeability fields obtained from the surrogate-based RML method: Case 1.

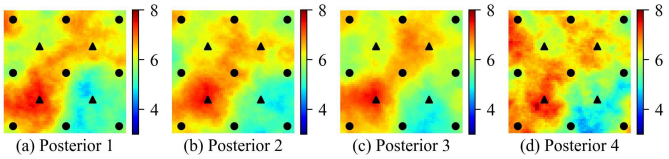

Fig. 10. Four posterior log-permeability fields obtained from the simulation-based RML method: Case 1.

5.1.2 History matching results

In this section, the surrogate model trained by 1,000 samples is used to replace the simulation in the history matching process, and the 50 time-steps (1,500 days) observations as historical data to be fit. As this study involves a synthetic model, we generated observed data by adding Gaussian noise with zero mean to the data simulated from the reference model. The standard deviation of the noise equals 5% of the "true" data. Upon completion of surrogate model training, we obtain explicit mathematical expressions mapping reservoir parameters to production data. Subsequently, with the surrogate model parameters fixed, and leveraging Pytorch's automatic differentiation functionality, we employ gradient-based optimization algorithms to update reservoir parameters. This approach efficiently optimized a series of perturbed objective functions within the RML framework to acquire the posterior distribution of reservoir parameters. The specific configuration for optimizing hyperparameters is detailed in Section 4.3.

Furthermore, for comparative purposes with the surrogate model-based RML sampling framework, we employ the Simultaneous Perturbation Stochastic Approximation algorithm based on numerical simulations to optimize the objective functions within the RML sampling framework. The hyperparameters for the Simultaneous Perturbation Stochastic Approximation algorithm are set as follows: initial learning rate of 2.5, initial perturbation step size of 0.05, and a maximum iteration step of 50. During each gradient estimation step, Simultaneous Perturbation Stochastic Approximation utilize five perturbation samples to calculate the corresponding estimated gradients differentially, followed by averaging to obtain the final gradient.

In this case, we perform dimension reduction based on 100 prior models. Utilizing the PCA method and setting the cumulative energy loss to 0.01, we reduce the dimensionality of the logarithmic permeability to 92. For both the surrogate-based RML method and the simulation-based RML method, we individually optimize 100 perturbed objective functions, ultimately yielding approximations of the posterior distribution of reservoir models by 100 posterior models.

Fig. 9 shows four posterior log-permeability fields obtained from the surrogate-based RML method. For comparison, Fig. 10 shows the posterior log-permeability fields obtained by the simulation-based RML method. Both methods basically capture the high-permeability and low-permeability regions of the reference model. In addition, the posterior models obtained by both methods still retain a certain uncertainty and have not occurred ensemble collapse phenomenon.

To further evaluate the inversion performance, we conduct numerical simulations using the posterior models obtained from both methods, as depicted in Figs. 11 and 12, respectively. Figs. 11 and 12 also present the production data from the reference model and the prior models for comparison. The total historical production period in this case spans 1,500 days, and compared to the simulation data from the prior models, the results of posterior models significantly reduce the uncertainty range. And both methods effectively fit the observed data.

Finally, we compare the computational efficiency of the surrogate-based RML method and the simulation-based RML method. The computational cost of the surrogate-based RML method primarily includes numerical simulations, surrogate model training, and numerical simulations of 100 posterior models. The training time of the surrogate model is approximately 5 minutes, with a total of 2,100 numerical simulation runs. In contrast, simulation-based RML method requires nearly 20,000 runs. It is evident that by introducing the surrogate model, computational resource consumption can be significantly reduced, which is highly advantageous for addressing the high-dimensional and computationally expensive sampling problem in automatic history matching.

5.2 Case 2: Heterogeneous waterflooding 3D reservoir model

In this case, a 3D reservoir model is used to estimate the proposed surrogate model. This model consists of 295,515 ($135\times199\times11$) grid cells. This model includes 50 prior models with different permeability fields. The reference permeability field is depicted in Fig. 13. In this case, the whole production period is 16 years and the interval for each time step is 30 days. The prediction objective of the surrogate model is the WOPR and WWPR.

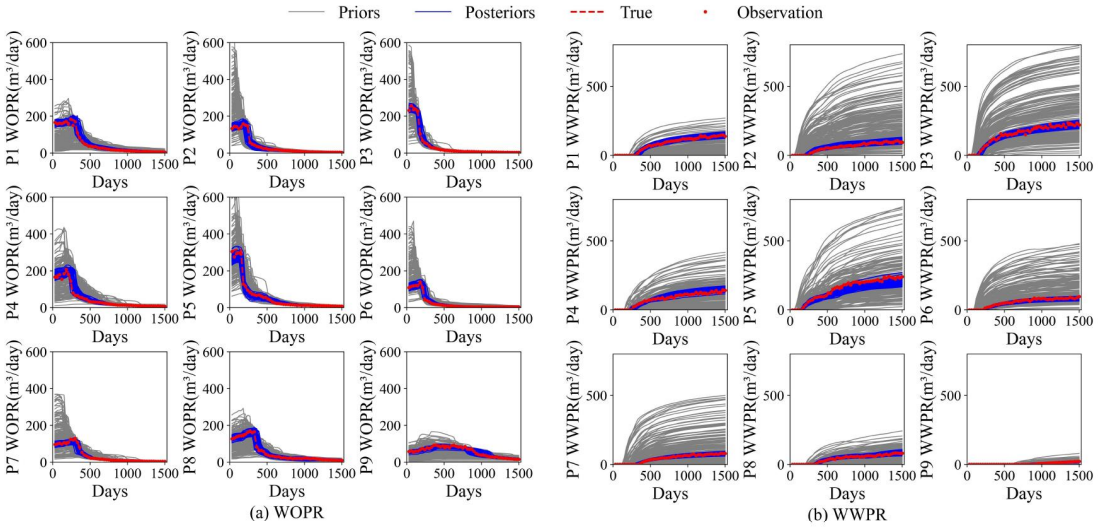

Fig. 11. Matching results of the posterior models obtained by surrogate-based RML method.

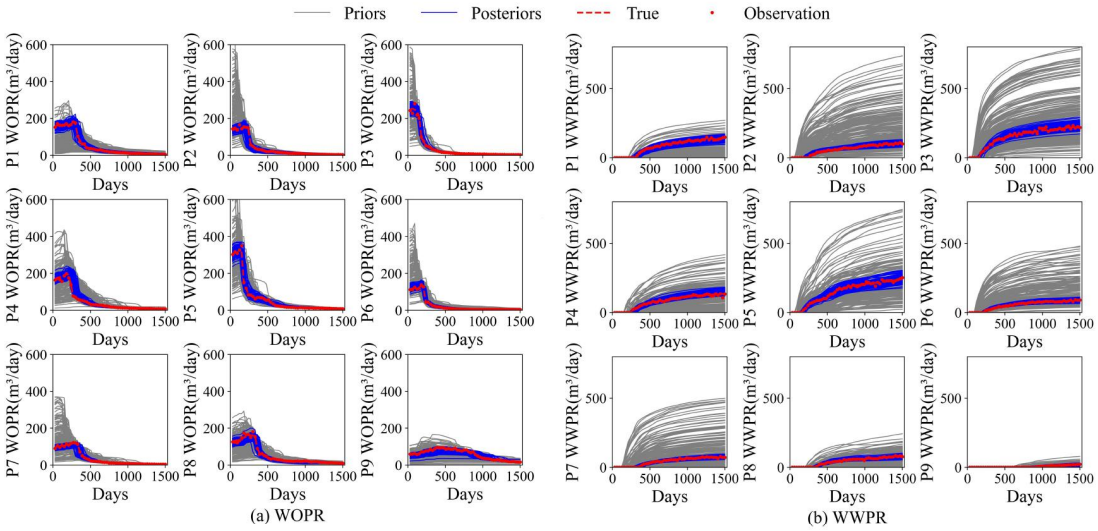

Fig. 12. Matching results of the posterior models obtained by simulation-based RML method.

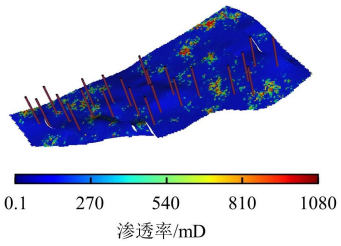

Fig. 13. Permeability of a random prior reservoir model: Case 2.

5.2.1 Prediction accuracy assessment

The case comprises a total of 50 prior models, and we utilize the PCA algorithm to reduce the dimensionality of the prior models, retaining 99% of the relative energy and reducing the parameters to 48 dimensions. Subsequently, we sample within the reduced-dimensional space to supplement insufficient samples. Given the large scale of this model, the runtime for a single numerical simulation is approximately 30 minutes.

Therefore, in this section, we sample only 1,000 models within the reduced-dimensional space, with 800 models allocated for the training set, 100 models for the validation set, and 100 models for the test set. The training epochs and batches for the surrogate model in this section were set to 100 and 16, respectively, with a training time of approximately 10 minutes. The surrogate model achieves an RMSE of 0.15656 and an R^2 value of 0.84824 on the test set.

5.2.2 History matching results

The proposed surrogate model trained by 800 samples is used to replace the simulation in the history matching process. We use the surrogate-based RML to optimize latent variables and finally obtain 100 posterior models. Fig. 14 present the mean and standard deviation of the permeability

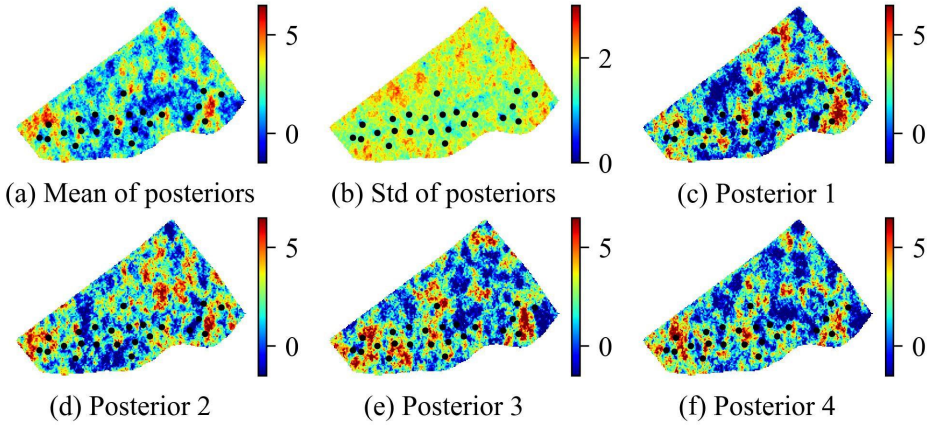

Fig. 14. Posterior log-permeability fields obtained from the surrogate-based RML method: Case 2.

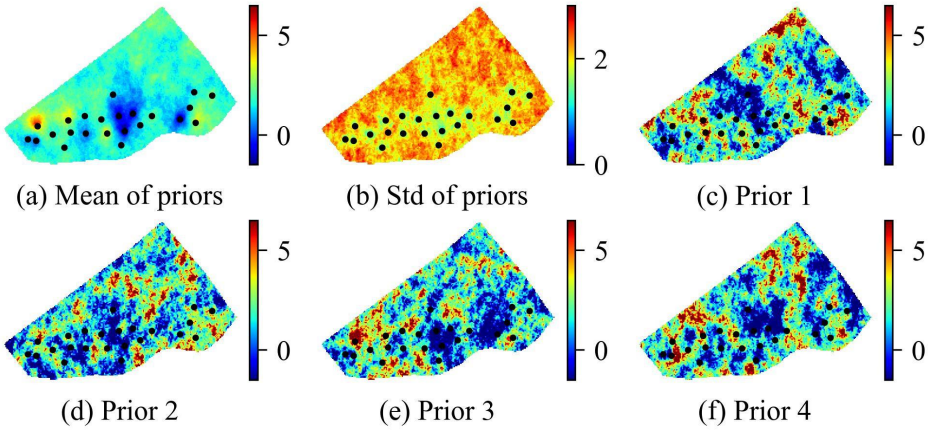

Fig. 15. Prior log-permeability fields: Case 2.

field in the second layer of the posterior models obtained using the surrogate-based RML method, as well as four randomly selected posterior permeability fields. For comparison with the posterior models, Fig. 15 depicts the mean and standard deviation of the permeability field in the second layer of the prior models, along with four randomly selected prior permeability fields. From an individual model perspective, the variation in permeability is not significant, and many characteristics of the prior model are preserved in the posterior model. In terms of the average permeability of the models, the prior model appears smoother than the posterior model, as the calibrated model highlights certain key features. Comparing the permeability standard deviations of the prior and posterior models, the standard deviation of permeability in the posterior model is significantly reduced compared to that in the prior model, with the reduction in uncertainty predominantly concentrated near the production and injection wells in the southern region. Additionally, it is evident that the posterior model still retains some level of uncertainty, indicating that the surrogate-based RML method has not experienced ensemble collapse and has produced a reasonable inversion result.

We simulate the posterior models obtained from the surrogate-based RML. To further illustrate the accuracy of the posterior reservoir models, we present the fitting results of daily oil and water production rates for selected wells and the entire reservoir, as shown in Fig. 16. Due to confidentiality requirements, the production data in this figure have been scaled. It is evident that the history matching results are highly successful, with each well exhibiting good agreement with the observed data while retaining a certain level of uncertainty. This indicates that the automatic history matching framework proposed in this study remains highly effective for large-scale reservoirs. Given the lengthy runtime of individual numerical simulations in this case, substituting numerical simulation processes with surrogate models will significantly enhance the efficiency of automatic history matching. This is highly beneficial for real-time decision-making in reservoir production management.

6. Conclusions

In this study, we introduced a novel approach, termed Convolution Recurrent Neural Network Surrogate Model-based Randomized Maximum Likelihood, for large-scale reservoir automatic history matching. Leveraging the flexibility and power of deep learning surrogate models within the framework of randomized maximum likelihood, our approach offers an efficient and accurate solution to the challenging problem of history matching in complex reservoir systems. Through extensive experimentation and analysis, we demonstrated the

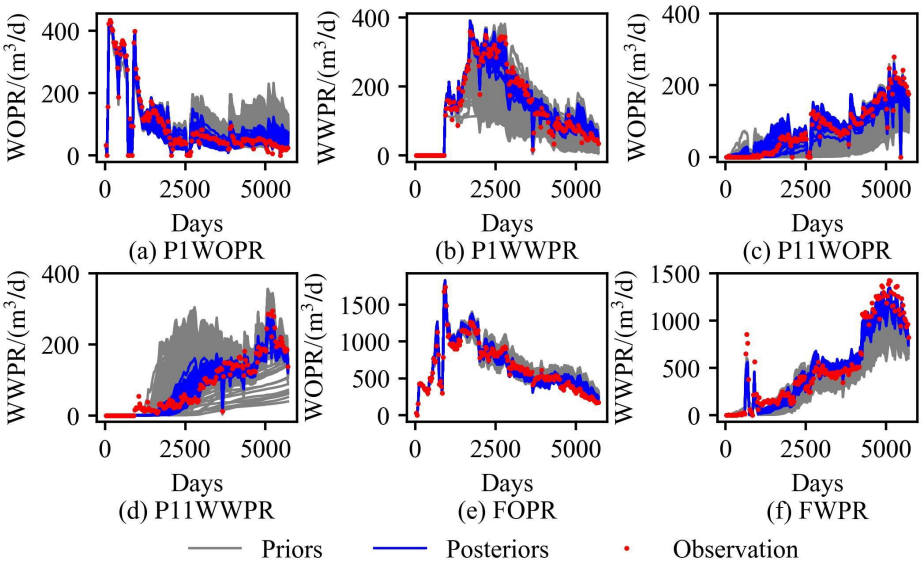

Fig. 16. Matching results of the posterior models obtained by surrogate-based RML method.

effectiveness of our approach in capturing the complex mapping relationship between heterogeneous reservoirs and production data. Furthermore, by employing surrogate models, we achieved substantial improvements in computational efficiency without sacrificing accuracy, thereby facilitating real-time decision-making in reservoir production management. In conclusion, the framework presented in this paper offers a powerful and practical solution for automatic history matching in large-scale reservoir systems. This approach is beneficial for reservoir management practices and contributes to more efficient and sustainable hydrocarbon recovery operations in the future.

Author information

The email addresses of the remaining authors of this paper are as follows:

zhangjinding@hotmail.com (J. Zhang);

zhangliming@upc.edu.cn (L. Zhang);

jsyanxia1989@163.com (X. Yan).

Acknowledgements

This work is supported by the National Natural Science Foundation of China under Grant 52325402, 52274057 and 52074340, the National Key R & D Program of China under Grant 2023YFB4104200, the Major Scientific and Technological Projects of CNOOC under Grant CCL2022RCPS0397RSN, 111 Project under Grant B08028.

Conflict of interest

The authors declare no competing interest.

References

Aanonsen, S. I., Noevdal, G., Oliver, D. S., et al. The ensemble kalman filter in reservoir engineering-a review. SPE Journal, 2009, 14(3): 393-412.

Asher, M. J., Croke, B. F., Jakeman, A. J., et al. A review of surrogate models and their application to groundwater modeling. Water Resources Research, 2015, 51(8): 5957-5973.

Cancelliere, M., Verga, F., Viberti, D. Benefits and limitations of assisted history matching. Paper SPE 146278 Presented at the SPE Offshore Europe Oil and Gas Conference and Exhibition, Aberdeen, UK, 6–8 September, 2011.

Chen, X., Zhang, K., Ji, Z. N., et al. Progress and challenges of integrated machine learning and traditional numerical algorithms: Taking reservoir numerical simulation as an example. mathematics, 2023, 11(21): 4418.

Emerick, A. A., Reynolds, A. C. EnKF-MCMC. Paper SPE 131375 Presented at the SPE EUROPEC/EAGE Annual Conference and Exhibition, Barcelona, Spain, 14–17 June, 2010.

Emerick, A. A., Reynolds, A. C. Ensemble smoother with multiple data assimilation. Computers and Geosciences. 2013, 55: 3-15.

Graves, A., Mohamed, A. R., Hinton, G. Speech recognition with deep recurrent neural networks. Paper Presented at IEEE international conference on acoustics, speech and signal processing, Vancouver, BC, Canada, 26-31 May, 2013.

Hamdi, H., Couckuyt, I., Sousa, M. C., et al. Gaussian Processes for history-matching: application to an unconventional gas reservoir. Computational Geosciences, 2017, 21(2): 267-287.

He, K. M., Zhang, X. Y., Ren, S. Q., et al. Deep residual learning for image recognition. Paper Presented at the IEEE Conference on Computer Vision and Pattern Recognition,

Las Vegas, Nevada, USA, 26 June - 1 July, 2016.

Hochreiter, S., Schmidhuber, J. Long short-term memory. Neural Computation, 1997, 9(8): 1735-1780.

Jafarpour, B., Goyal, V. K., McLaughlin, D. B., et al. Compressed history matching: exploiting transform-domain sparsity for regularization of nonlinear dynamic data integration problems. Mathematical Geosciences, 2010, 42: 1-27.

Kingma, D. P., Ba, J. Adam: A method for stochastic optimization. arXiv preprint, arXiv: 1412.6980, 2014.

Kitanidis, P. K. Parameter uncertainty in estimation of spatial functions: Bayesian analysis. Water Resources Research, 1986, 22(4): 499-507.

LeCun, Y., Bengio, Y., Hinton, G. Deep learning. Nature, 2015, 521(7553): 436-444.

Li, X., Reynolds, A. C. A gaussian mixture model as a proposal distribution for efficient markov-chain monte carlo characterization of uncertainty in reservoir description and forecasting. SPE Journal, 2020, 25(1): 1-36.

Liu, N., Oliver, D. S. Evaluation of monte carlo methods for assessing uncertainty. SPE Journal, 2003, 8(2): 188-195.

Lu, P., Horne, R. N. A multiresolution approach to reservoir parameter estimation using wavelet analysis. Paper SPE 62985 Presented at SPE Annual Technical Conference and Exhibition, Dallas, Texas, 1–4 October, 2000.

Ma, X. P., Zhang, K., Wang, J., et al. An efficient spatial-temporal convolution recurrent neural network surrogate model for history matching. SPE Journal, 2022a, 27(2): 1160-1175.

Ma, X. P., Zhang, K., Zhang, J. D., et al. A novel hybrid recurrent convolutional network for surrogate modeling of history matching and uncertainty quantification. Journal of Petroleum Science and engineering, 2022b, 210: 110109.

Mo, S. X., Zhu, Y. H., Zabaras, N., et al. Deep convolutional encoder-decoder networks for uncertainty quantification of dynamic multiphase flow in heterogeneous media. Water Resources Research, 2019, 55 (1): 703-728.

Oliver, D. S., He, N., Reynolds, A. C. Conditioning permeability fields to pressure data. Presented at the ECMOR V-5th European Conference on the Mathematics of Oil Recovery, 1996.

Oliver, D. S., Reynolds, A. C. L, N. Inverse theory for petroleum reservoir characterization and history matching. Cambridge, UK: Cambridge University Press, 2008.

Paszke, A., Gross, S., Massa, F., et al. Pytorch: An imperative style, high-performance deep learning library. Advances in Neural Information Processing Systems, 2019, 32.

Remy, N., Boucher, A., Wu, J. Applied geostatistics with SGeMS: A user's guide. Cambridge, UK: Cambridge University Press, 2009.

Reynolds, A. C., He, N., Chu, L., et al. Reparameterization techniques for generating reservoir descriptions conditioned to variograms and well-test pressure data. SPE Journal, 1996, 1(4): 413-426.

Sarma, P., Durlofsky, L. J., Aziz, K., et al. Efficient real-time reservoir management using adjoint-based optimal control and model updating. Computational Geosciences, 2006, 10(1): 3-36.

Tang, M., Liu, Y., Durlofsky, L. J. A deep-learning-based surrogate model for data assimilation in dynamic subsurface flow problems. Journal of Computational Physics, 2020, 413: 109456.

Van Leeuwen, P. J., Evensen, G. Data assimilation and inverse methods in terms of a probabilistic formulation. Monthly weather review, 1996, 124(12): 2898-2913.

Wantawin, M., Yu, W., Dachanuwattana, S., et al. An iterative response-surface methodology by use of high-degree-polynomial proxy models for integrated history matching and probabilistic forecasting applied to shale-gas reservoirs. SPE Journal, 2017, 22(6): 2012-2031.

Xiao, C., Leeuwenburgh, O., Lin, H. X., et al. Conditioning of deep-learning surrogate models to image data with application to reservoir characterization. Knowledge-Based Systems, 2021.

Yan, L., Zhou, T. An adaptive surrogate modeling based on deep neural networks for large-scale bayesian inverse problems. Communications in Computational Physics, 2020, 28 (5): 2180-2205.

Yu, W., Tripoppoom, S., Sepehrnoori, K., et al. An automatic history-matching workflow for unconventional reservoirs coupling MCMC and non-intrusive EDFM methods. Paper SPE 191473 Presented at the SPE Annual Technical Conference and Exhibition, Dallas, Texas, USA, 24–26 September, 2018.

Zhang, J. J., Lin, G., Li, W. X., et al. An iterative local updating ensemble smoother for estimation and uncertainty assessment of hydrologic model parameters with multimodal distributions. Water Resources Research, 2018, 54(3): 1716-1733.

Zhong, Z., Sun, A. Y., Wang, Y., et al. Predicting field production rates for waterflooding using a machine learning-based proxy model. Journal of Petroleum Science and Engineering, 2020, 194: 107574.

Zhu, Y. H., Zabaras, N. Bayesian deep convolutional encoder-decoder networks for surrogate modeling and uncertainty quantification. Journal of Computational Physics, 2018, 366: 415-447.