Journal of Petroleum Exploration and Production Technology (2025) 15:32
https://doi.org/10.1007/s13202-025-01937-5

ORIGINAL PAPER - PRODUCTION ENGINEERING

Learning-based pattern-data-driven forecast approach for predicting future well responses

Yeongju Kim[1] · Baehyun Min[2] · Alexander Sun[3] · Bo Ren[3] · Hoonyoung Jeong[1,4]

Received: 22 April 2024 / Accepted: 14 January 2025 / Published online: 7 February 2025
© The Author(s) 2025

Abstract

Forecasting well responses, such as flow rates and pressures, is crucial for effective reservoir management and investment decision-making in the development of subsurface reservoir resources. Recently, data-driven forecast methods, such as data-space inversion (DSI) and a learning-based data-driven forecast approach (LDFA), have been introduced to mitigate the computational cost and geological constraint issues of history-matching methods. However, DSI and LDFA have extrapolation, conditioning, and prediction variance issues. In this study, we propose two simpler alternatives, a learning-based pattern-data-driven forecast approach (LPFA) and an ensemble conditioning step (ECS), to resolve the issues associated with DSI and LDFA. LPFA mitigates the extrapolation issue by scaling well responses for each instance using the mean and variance of an observation period. ECS addresses the conditioning and prediction variance issues of LDFA and LPFA by combining predictions of multiple learning models and screening out predictions that do not sufficiently honor observed data. The prediction performances of DSI, LDFA, LPFA, and ECS were compared using two benchmark models: Brugge and Olympus models. Among these methods, LPFA provided the most accurate predictions for future well responses and reasonable uncertainty intervals, achieving an error of 2.3%. Even when predicting well responses outside the range of prior data, LPFA maintained a prediction error of 4.8%, unlike the other methods whose performance significantly declined. ECS improved the prediction accuracy by 1–2% and reduced the uncertainty in the predictions of future well responses by approximately 50%. Our approaches are generic and can be integrated with other data-driven forecast methods to enhance prediction performance.

Keywords Data-driven forecast · Learning-based forecast · History matching · Machine learning

Abbreviations

DSI	Data-space inversion
ECS	Ensemble conditioning step
ES-MDA	Ensemble smoother with multiple data assimilation
IES	Iterative ensemble smoother
ML	Machine learning
LDFA	Learning-based data-driven forecast approach
LPFA	Learning-based pattern-driven forecast approach
OWC	Oil-water contact
PCA	Principal component analysis
PDF	Probability density function
RML	Randomized maximum likelihood
SGS	Sequential Gaussian simulation
SIS	Sequential indicator simulation

✉ Hoonyoung Jeong
hoonyoung.jeong@snu.ac.kr

[1] Department of Energy Systems Engineering, Seoul National University, 1, Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea

[2] Department of Climate and Energy Systems Engineering, Ewha Womans University, 52 Ewhayeodae-gil, Seodaemun-gu, Seoul 03760, Republic of Korea

[3] Bureau of Economic Geology, Jackson School of Geosciences, The University of Texas at Austin, Austin, TX 78758, United States of America

[4] Research Institute of Energy and Resources, Seoul National University, 1, Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea

Latin Letters

$\widetilde{C}$	Covariance matrix of posterior data
C_e	Covariance matrix of ϵ
C_M	Covariance matrix of prior reservoir model parameters
C_{pd}	Covariance matrix of prior data
C_η	Covariance matrix of η
d	Vector of well responses
$\widetilde{d}$	Mean vector of posterior data
d_f	Vector of well responses at forecast times
$\widehat{d}_f$	Prediction results
d_{in}	Matrix of data vector of all prior reservoir models for observation times
$d_{in,\,scaled}$	Scaled matrix of d_{in} (dimensionless)
$d_{in,\,standardized}$	Standardized matrix of d_{in} (dimensionless)
d_{obs}	Vector of observed data
d_{out}	Matrix of data vector of all prior reservoir models for forecast times
$d_{out,\,scaled}$	Scaled matrix of d_{out} (dimensionless)
$d_{out,\,standardized}$	Standardized matrix of d_{out} (dimensionless)
d_{prior}	Mean vector of prior data
d_{pca}	Matrix of new data realizations using PCA
d_{t_k}	Vector of well responses at time step t_k
$\bar{d}_{\tau\,i}$	Mean vector of prior data at $\tau_{\,i}$
$\bar{d}_{\tau\,,l}$	Mean vector of prior data during observation times in l-th prior reservoir model.
E_{N_l}	Cumulative energy ratio of N_l eigenvectors (dimensionless)
f_1	Function of the reparameterization step
f_1^{-1}	Back-transform function of f_1
$f_\eta\,(\bullet)$	Function of reparameterization from η to d_f
$f_\xi\,(\bullet)$	Function of reparameterization from ξ to d_f
$g\,(\bullet)$	Forward reservoir simulation
H	Matrix for extracting well response that correspond to observation times
$h_T\,(\bullet)$	Histogram transformation
J	Objective function (dimensionless)
M	Matrix of prior reservoir model parameters
m	Vector of reservoir model parameters
m_{prior}	Vector of prior estimates of the model parameters
N_e	Number of prior reservoir models (dimensionless)
N_f	Number of forecast processes (dimensionless)
N_{in}	Number of chosen eigenvectors of d_{in} (dimensionless)
N_l	Number of chosen eigenvectors (dimensionless)
N_m	Number of learning models (dimensionless)
N_{out}	Number of chosen eigenvectors of d_{out} (dimensionless)
N_t	Number of forecast times (dimensionless)
N_τ	Number of observation times (dimensionless)
$p\,(\bullet)$	Probability density function (dimensionless)
$t_1,\,t_2,\,\ldots,\,t_{N_t}$	Forecast times (day)
$\widehat{x}_f$	Output vector projected on the dimension-reduced space of the training output data
x_{obs}	Vector of the observations projected on a dimension-reduced space of the training input data

Greek Letters

ϵ	Vector of measurement and modeling error (%)
ϵ	Relative error (dimensionless)
$\epsilon_{criterion}$	Criterion of relative error (dimensionless)
η	Reparameterized vector by histogram transformation
$\eta_{\,prior}$	Mean vector of η
λ	Eigenvalue
$\mathcal{L}\,(\bullet)$	Learning model
Φ	Basis matrix of PCA
$\sigma_{\,t_i}$	Standard deviation vector of prior data at t_i
$\sigma_{\,\tau\,i}$	Standard deviation vector of prior data at $\tau_{\,i}$
$\sigma_{\,\tau\,,l}$	Standard deviation vector of prior data during observation times in *l*-th prior reservoir model.
$\tau_1,\,\tau_2,\,\ldots,\,\tau_{N_\tau}$	Observation times (day)
ξ	Reparameterized vector by PCA

Introduction

A well response refers to the behavior of a producer or injector, such as flow rates, bottom-hole pressure, and cumulative flow rates, in response to various factors like reservoir properties and operational conditions. Forecasting well responses is crucial for effective reservoir management and investment decision-making in the development of subsurface reservoir resources, including oil and gas production, geothermal energy production, and geological carbon storage. History-matching methods have been employed to achieve reliable forecasts of well responses. In conventional history-matching methods, prior reservoir models are calibrated to match simulated data to observed data. First, an ensemble of prior models is generated based on well log, core, and seismic data. In the calibration step, reservoir model parameters are updated to minimize the objective function using optimization algorithms. Various optimization algorithms have been developed, including gradient-based methods (Chen et al. 1974; Chavent, Dupuy, et al. 1975; Li et al. 2003; Gao et al. 2007; Bertolini and Schiozer 2011) and ensemble-based methods (Chen and Oliver 2013; Luo et al. 2015; Kim et al. 2018; Jung et al. 2020; Kang et al. 2020). For example, a Quasi-Newton method, an iterative ensemble smoother (IES), and an ensemble smoother with multiple data assimilation (ES-MDA) are commonly used. Future well responses are then predicted by numerically simulating the calibrated reservoir models (Rwechungura et al. 2011; Liu et al. 2023a). However, hundreds of forward reservoir simulation runs may be required in the calibration step of history-matching methods. Additional forward reservoir simulation runs may be needed to calibrate other prior reservoir models representing reservoir uncertainty. Thus, history-matching methods are resource-intensive and challenging, especially when time is limited.

Additionally, it is difficult to maintain geological constraints imposed on prior reservoir models in the calibration step of history-matching methods. Objective functions of history-matching methods typically consist of two misfits: the misfit between the simulated and observed data and between the prior and posterior model parameters. If the misfit between the simulated and observed data is excessively weighted in the objective function, the posterior model parameters may not satisfy the geological constraints. Thus, history-matching methods have a trade-off issue for misfits.

Sun and Durlofsky (2017) proposed the data-space inversion (DSI) method, which forecasts well responses rapidly without calibrating the prior reservoir models. In this method, well responses are treated as random data variables with a prior probability density function (PDF). The prior PDF is estimated from simulated data of the prior reservoir models. Using Bayes' theorem, DSI then estimates the posterior PDF of the well responses that honor the observed data. Future well responses are predicted by sampling from the posterior PDF once it is established. If the prior PDF is non-Gaussian, a transformation is required, involving a mapping operation and principal component analysis (PCA). After the transformation, DSI employs the randomized maximum likelihood (RML) method to sample future well responses from the posterior PDF. Sun and Durlofsky (2017) originally applied a simple mapping operation that shifts, compresses, and stretches time series for transforming non-Gaussian data into Gaussian form. However, employing a simple mapping operation is difficult when well operating conditions change sharply, as the transformations need clear patterns of well responses. To address this, Sun et al. (2017) implemented a histogram transformation instead.

Other researchers have extended DSI for various applications. Jiang et al. (2020); Liu et al. (2023b) treated well operating conditions as random variables and applied DSI to optimize the well operating conditions. Liu et al. (2021) employed DSI for forecasting spatial data, such as saturation and pressure profiles. Lima et al. (2020) reduced the computational cost of DSI using an ensemble smoother instead of gradient-based optimization methods.

Beyond DSI, machine learning (ML) has also been widely adopted to forecast well responses (Liang et al. 2023; Bigdeli and Delshad 2024). Some studies have focused on improving efficiency in history matching using machine-learning (Kim et al. 2020; Jiang and Durlofsky 2021; Jo et al. 2022; Liang et al. 2023). Others have introduced forecast approaches that do not rely on history matching, instead solely based on a data-driven approach (Nwachukwu et al. 2018; Jeong et al. 2018; Lee et al. 2019; Zhong et al. 2019, 2020, 2021; Wang et al. 2021; Chahar et al. 2022; Xiao et al. 2022, 2023a, b). Furthermore, efforts have been made to integrate DSI with ML for enhanced performance (Xiao et al. 2022, 2023a, b).

Jeong et al. (2018) proposed a learning-based data-driven forecast approach (LDFA) that can quickly predict future well responses. In LDFA, prior data, well responses of prior reservoir models, are used as training data. The training data are encoded and decoded using PCA to reduce the computational cost of the training process. LDFA builds a learning model reproducing the data-driven relationship between historical and forecast data from an ensemble of prior reservoir models. Then, LDFA forecasts future well responses by inputting observed data into the learning models. LDFA is computationally efficient and easier to implement than DSI because it employs well-established ML techniques, such as artificial neural networks and support vector regression.

However, neither DSI nor LDFA provides satisfactory forecast results for observed data that fall outside the range of prior data (i.e., extrapolating outside the range of prior data). This extrapolation issue arises from PCA and machine learning techniques. PCA is a common dimension reduction technique suited for data interpolation rather than extrapolation. Extrapolating with PCA can be problematic because PCA assumes the primary structure captured by the principal components will hold true for unseen data (Jolliffe 2002). For this reason, DSI has the extrapolation issue because PCA is used in the histogram transformation that is applied to input data in DSI. Machine learning techniques that are used to predict future well responses in LDFA are also often challenged in extrapolation tasks, as they are primarily designed for interpolation within the range of prior data (Goodfellow et al. 2016).

LDFA also has issues of conditioning and prediction variance. The input data of LDFA are well responses from the initial time to the last observation time, whereas the output data of LDFA are well responses from the initial time to a future time. However, LDFA predictions may not be reliable unless the predictions are conditioned to the observed data, whereas DSI has an iterative process to honor the observed data. Furthermore, the predictions of learning models may have high variances even when the same training data are used because the learnable variables, such as weights and biases, are randomly initialized (Glorot and Bengio 2010; Scardapane and Wang 2017). The random initialization of the learnable variables may lead to learning models that return significantly different predictions across multiple instances of model training (Ahmed and Lofstead 2022).

This study proposes two simple approaches to mitigate the extrapolation, conditioning, and prediction variance issues of DSI and LDFA. The first approach involves the development of a learning-based pattern-driven forecast approach (LPFA) to solve the extrapolation issue by adding a new scaling method to LDFA. The new scaling method scales prior data for each instance (realization) using its mean and standard deviation during an observation period. The new scaling method is straightforward but can significantly mitigate the extrapolation issue of DSI and LDFA. The second approach introduces an ensemble conditioning step (ECS) to address both the conditioning and prediction variance issues in DSI and LDFA. ECS finds and selects predictions honoring observed data in multiple learning models. ECS can improve both accuracy and variance of predictions.

In the following section, the formulations and procedures of DSI, LDFA, LPFA, and ECS are presented. Then, the performances of DSI, LDFA, and LPFA, with and without ECS, are compared for the well-known benchmark models, the Brugge and Olympus models. In both benchmark models, LPFA provided the most accurate predictions for future well responses even when the observed data were far from the prior data. ECS improved the prediction performance of LDFA and LPFA.

Methodology

This section outlines the procedures for the two data-driven forecast methods (DSI and LDFA) and introduces the two proposed methods (LPFA and ECS). The "Preliminaries for Data-driven Forecast Approaches" subsection provides the necessary background information and common notations required for understanding DSI and LDFA. The "Data Space Inversion" and "Learning-based Data-driven Forecast Approach" subsections detail the procedures and formulations of DSI and LDFA. Finally, the "Learning-based Pattern-driven Forecast Approach" and "Ensemble Conditioning Step" subsections explain how the proposed methods address extrapolation, conditioning, and prediction variance issues.

- Preliminaries for Data-driven Forecast Approaches.

Let m denote a vector of reservoir model parameters, such as permeability and porosity. To account for the uncertainty of the reservoir model parameters, an ensemble of N_e reservoir models, $M = [m_1, m_2, \ldots, m_{N_e}]$, is generated. M is referred to as the prior reservoir models. The prior data, the well responses of the prior reservoir models, are obtained from forward reservoir simulations. The well responses consist of flow rates, bottom-hole pressure, or cumulative flow rates. This procedure can be expressed as Eq. (1).

$$d_{f,l} = g\left(m_l\right),\ l = 1,2,\ \ldots,\ N_e, \tag{1}$$

where $g(\bullet)$ returns a vector of forward reservoir simulation results for the forecast times, and $d_{f,l}$ is a vector of well responses at the forecast times for the l-th ensemble model. d_f can be expressed as Eq. (2).

$$d_f = \left[d_{t_1}^T,\ d_{t_2}^T,\ \ldots,\ d_{t_{N_t}}^T\right]^T \tag{2}$$

where t is the forecast time, N_t is the total number of the forecast times (t_1, t_2,..., t_{N_t}), and d_{t_k} is a well response at time step t_k.

Let d_{obs} denote the observed data, which is a vector of observations. d_{obs} can be expressed as Eq. (3).

$$d_{obs} = \left[d_{\tau_1}^T, d_{\tau_2}^T, \ldots, d_{\tau_{N_\tau}}^T \right]^T, \quad \{\tau_1, \tau_2, \ldots, \tau_{N_\tau}\} \subseteq \{t_1, t_2, \ldots, t_{N_t}\} \tag{3}$$

where τ and N_τ are the observation time and the total number of observation times ($\tau_1, \tau_2, \ldots, \tau_{N_\tau}$), respectively. The set of the observation times is a subset of the forecast times.

- Data Space Inversion.

DSI generates multiple future well responses conditioned to observed data within a Bayesian framework (Sun and Durlofsky 2017). In DSI, well response is treated as random data variable with a prior PDF. This prior PDF is estimated from the prior data. Once the posterior distribution of data variable given observed data is calculated, future well responses are generated by directly sampling from the posterior PDF. Figure 1 shows the procedure of DSI.

According to Bayes' theorem, the posterior PDF of well response, $p(d_f|d_{obs})$, is expressed as Eq. (4) (Bayes 1763).

$$p(d_f|d_{obs}) = \frac{p(d_{obs}|d_f)\,p(d_f)}{p(d_{obs})} \propto p(d_{obs}|d_f)\,p(d_f), \tag{4}$$

where $p(d_f)$ is the prior PDF of well response, and $p(d_{obs}|d_f)$ is the likelihood function. As d_{obs} is a vector of well responses at the observation times, the d_{obs} and d_f are related as Eq. (5).

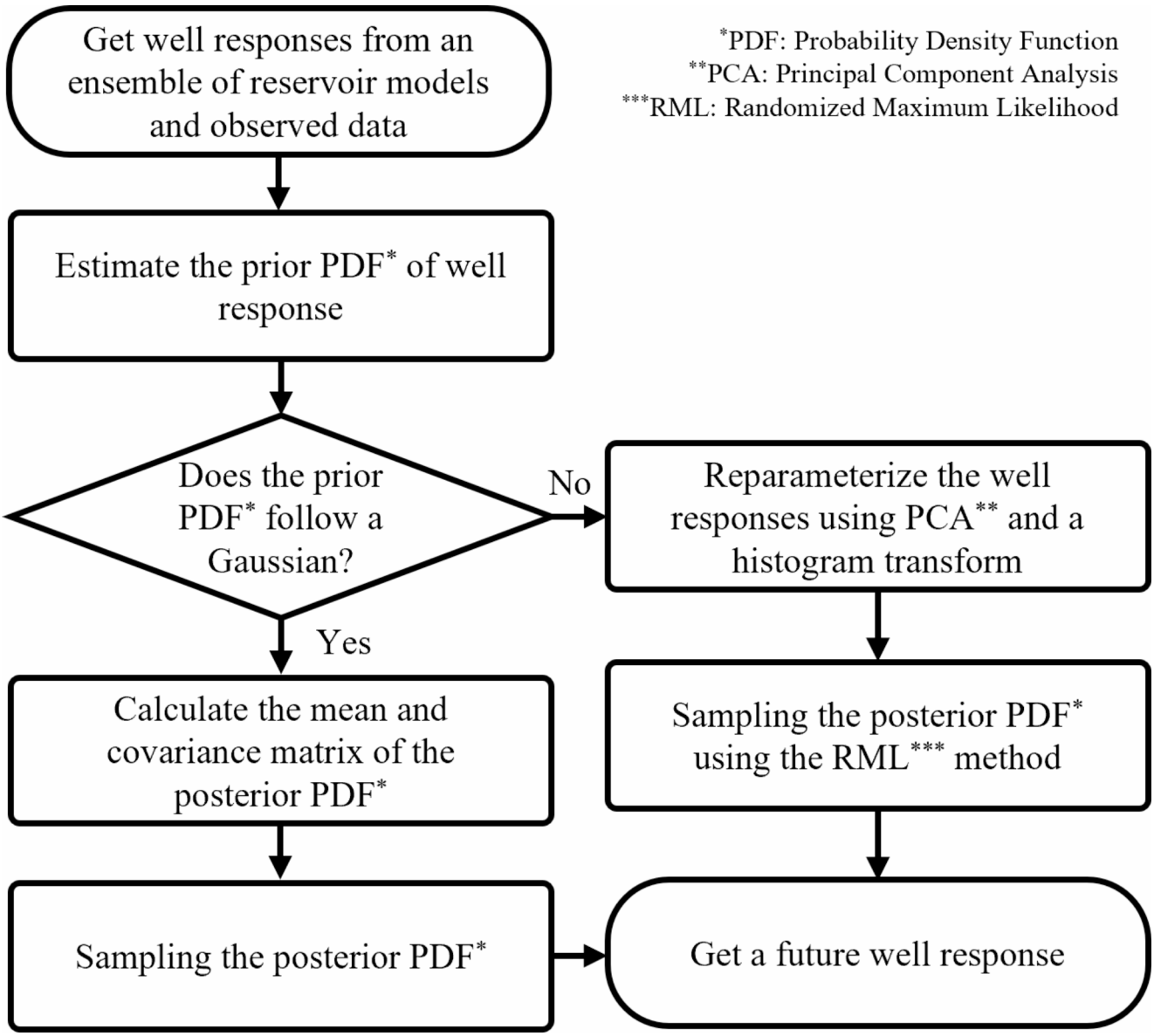


Fig. 1 Flowchart of data space inversion (DSI)

$$d_{obs} = Hd_f + \epsilon\,. \tag{5}$$

where H is a matrix for extracting simulation results that correspond to the observation times, and ϵ is a vector of the measurement and modeling errors. ϵ is assumed to follow a Gaussian distribution with a zero mean and a covariance matrix C_e.

As the error vector ϵ follows a Gaussian distribution with zero mean and a covariance matrix C_e, $p\left(d_{obs}|d_f\right)$ is expressed as Eq. (6).

$$\begin{aligned} &p\left(d_{obs}|d_f\right) = p\left(\int\ = d_{obs} - Hd_f\right) \propto \\ &\exp\left(-\frac{1}{2}\left(Hd_f - d_{obs}\right)^T C_e^{-1}\left(Hd_f - d_{obs}\right)\right) \end{aligned} \tag{6}$$

If $p\left(d_f\right)$, follows a multivariate Gaussian distribution with a mean vector d_{prior} and a covariance matrix C_{pd}, the prior PDF is expressed as Eq. (7).

$$p\left(d_f\right) \propto \exp\left(-\frac{1}{2}(d_f - d_{prior})^T C_{pd}^{-1}\left(d_f - d_{prior}\right)\right). \tag{7}$$

The mean vector, d_{prior}, and the covariance matrix, C_{pd}, are calculated from the prior data as Eqs. (8)-(9).

$$d_{prior} = \frac{1}{N_e}\sum\nolimits_{l=1}^{N_e} d_{f,\,l} \tag{8}$$

$$C_{pd} = \frac{1}{N_e - 1}\sum\nolimits_{l=1}^{N_e}\left(d_{f,\,l} - d_{prior}\right)\left(d_{f,\,l} - d_{prior}\right)^T. \tag{9}$$

Combining Eqs. (4), (6), and (7), the posterior PDF of well response, $p\left(d_f|d_{obs}\right)$, is expressed as Eq. (10).

$$\begin{aligned} p\left(d_f|d_{obs}\right) \propto exp\Bigg(&-\frac{1}{2}\left(Hd_f - d_{obs}\right)^T C_e^{-1}\left(Hd_f - d_{obs}\right) \\ &-\frac{1}{2}\left(d_f - d_{prior}\right)^T C_{pd}^{-1}\left(d_f - d_{prior}\right)\Bigg). \end{aligned} \tag{10}$$

Then, the mean and covariance matrix of the posterior PDF are expressed as Eqs. (11)-(12).

$$\tilde{d} = d_{prior} + \left(H^T C_e^{-1} H + C_{pd}^{-1}\right)^{-1} H^T C_e^{-1}\left(d_{obs} - Hd_{prior}\right), \tag{11}$$

$$\tilde{C} = \left(H^T C_e^{-1} H + C_{pd}^{-1}\right)^{-1}, \tag{12}$$

where $\tilde{d}$ and $\tilde{C}$ are the mean vector and covariance matrix of the posterior PDF, respectively. Future well responses are predicted by sampling the posterior PDF.

If $p\left(d_f\right)$ does not follow a multivariate Gaussian distribution, the data variables d_f need to be reparameterized to new variables η that follow a multivariate Gaussian distribution. Then, d_f can be expressed as Eq. (13).

$$d_f = f_\eta\left(\eta\right), \tag{13}$$

where f_η is the function of reparameterization from η to d_f.

Let η follow a multivariate Gaussian distribution with mean $\eta_{\,prior}$ and covariance C_η. Then, the conditional distribution of η given d_{obs} can be expressed as Eq. (14).

$$\begin{aligned} &p\left(\eta|d_{obs}\right) \propto p\left(d_{obs}|\eta\right) p\left(\eta\right) \propto exp \\ &\Bigg(-\frac{1}{2}\left(Hf\left(\eta\right) - d_{obs}\right)^T C_e^{-1}\left(Hf\left(\eta\right) - d_{obs}\right) \\ &-\frac{1}{2}\left(\eta - \eta_{prior}\right)^T C_\eta^{-1}\left(\eta - \eta_{prior}\right)\Bigg) \end{aligned} \tag{14}$$

The data variable is reparameterized using PCA and a histogram transformation. Let ξ denote the reparameterized vector representing d_f. Then, d_f can be expressed using ξ as shown in Eq. (15).

$$d_f \approx\ f_\xi\left(\xi\right) = h_T\left(d_{pca}\right) = h_T\left(\Phi\,\xi\ + d_{prior}\right), \tag{15}$$

where f_ξ is the function of reparameterization from ξ to d_f, h_T is the histogram transformation, Φ is the basis matrix of PCA, and d_{pca} is a matrix of the new data realizations using PCA. Then, the posterior distribution of new parameters ξ given d_{obs} can be expressed as Eq. (16).

$$\begin{aligned} p\left(\xi\,|d_{obs}\right) \propto\ &\exp\Bigg(-\frac{1}{2}\left(Hf\left(\xi\right) - d_{obs}\right)^T \\ &C_e^{-1}\left(Hf\left(\xi\right) - d_{obs}\right) - \frac{1}{2}\xi^{\,T}\xi\Bigg) \end{aligned} \tag{16}$$

For sampling the posterior PDF, the RML method is used. The RML method finds ξ that maximizes the posterior PDF, $p\left(\xi\,|d_{obs}\right)$. The procedure of the RML method is defined in Eq. (17).

$$\begin{aligned} \xi_{\,rml} = \arg\min_{\xi}\Big[&\left(Hf\left(\xi\right) - d^*_{obs}\right)^T C_e^{-1}\left(Hf\left(\xi\right) - d^*_{obs}\right) \\ &+\left(\xi\ - \xi^{\,*}\right)^T\left(\xi\ - \xi^{\,*}\right)\Big] \end{aligned} \tag{17}$$

,

where $\xi_{\,rml}$ is the result of the RML method, d^*_{obs} represents perturbed observations from a Gaussian distribution $N\left(d_{obs},\ C_e\right)$, and $\xi^{\,*}$ represents variables sampled from the standard normal distribution $N\left(0,\ I\right)$.

In conventional history-matching methods, an objective function is formulated to quantify the discrepancy between

the simulated and observed data and between the prior and posterior model parameters as Eq. (18) (Oliver et al. 2008).

$$J(m) = \frac{1}{2}(d_{obs} - H \cdot g(m))^T C_e^{-1}(d_{obs} - H \cdot g(m)) + \frac{1}{2}(m - m_{prior})^T C_M^{-1}(m - m_{prior}) \quad (18)$$

where $J(m)$ is the objective function, m_{prior} is the prior estimate of the model parameters and C_M is the covariance matrix of the prior model parameters. Equation (10) is a function of observable variables such as well responses while Eq. (18) is a function of model parameters such as permeability and porosity. If the number of prior models is 100, only 100 simulation runs are required to predict future well responses in the data-driven forecast approaches including DSI, LDFA, and LPFA. Even though additional observations are obtained, the data-driven forecast approaches do not need additional simulation runs.

On the contrary, ES-MDA, which is one of the commonly used ensemble-based history matching methods, requires more simulation runs than the data-driven forecast approaches. In the ES-MDA procedure, model parameters are typically assimilated four times consecutively (Emerick and Reynolds 2013). If the assimilations are conducted four times, ES-MDA requires $(4+1) \times 100 = 500$ simulation runs to predict future well responses. If additional observations are obtained, ES-MDA needs another 500 simulation runs. Gradient-based methods require more simulation runs to compute the gradient vector of the objective function in the iterations. In this regard, the data-driven forecast approaches are more computationally efficient than the conventional history matching methods when predicting future well responses.

Compared with history matching methods, one drawback of the data-driven forecast approaches is that they do not provide calibrated model parameters such as porosity and permeability. The data-driven forecast approaches predict future well responses without calibrating prior models, relying instead on the relationship between prior data that are obtained by conducting reservoir simulation in prior models. However, this drawback can be advantageous in terms of addressing the ill-posedness of inversion problems. Maintaining geological constraints imposed on prior models during the calibration step of history matching methods is challenging, even when using a prior term in the objective function, as given in Eq. (18). The data-driven forecast approaches, by not requiring calibration of model parameters, can avoid the ill-posed problems inherent in conventional history matching methods.

- Learning-based Data-driven Forecast Approach.

LDFA generates multiple future well responses using an ML model. In LDFA, the learning model is trained to reproduce the data-driven relationship between well responses before and after the observation time from the prior data. Once the data-driven relationship is trained, the learning model generates future well responses from the observed data. The learning model takes well responses during the observation period as input data and predicts well responses for the entire period. As the dimensions of the input and output data are large, the computational cost of training becomes high. Therefore, in LDFA, the dimensions of the input and output data are reduced using PCA to save the computational cost. By reducing the data dimensions, the learning model can learn the data-driven relationship using a simple feed-forward network with one hidden layer. Due to its simple structure, LDFA has the advantage of obtaining prediction results quickly. Figure 2 shows the procedure of LDFA.

The prior data are used as training data for a machine learning model. The training input and output data is described in Eqs. (19)-(20).

$$d_{out} = [g(m_1), \ldots, g(m_{N_e})] = \begin{bmatrix} d_{t_1,\,1} & \cdots & d_{t_1,\,N_e} \\ \vdots & \ddots & \vdots \\ d_{t_{N_t},\,1} & \cdots & d_{t_{N_t},\,N_e} \end{bmatrix}, \quad (19)$$

$$d_{in} = H d_{out} = \begin{bmatrix} d_{\tau_1,\,1} & \cdots & d_{\tau_1,\,N_e} \\ \vdots & \ddots & \vdots \\ d_{\tau_{N_\tau},\,1} & \cdots & d_{\tau_{N_\tau},\,N_e} \end{bmatrix} \quad (20)$$

where d_{out} and d_{in} represent the training input and output data, respectively.

If the dimensions of the training input and output data, $N_\tau \times N_e$ and $N_t \times N_e$, are large, it is computationally expensive to train a learning model. Thus, PCA reduces the dimensions of the training input and output data. PCA calculates the eigenvectors (the principal direction of a covariance matrix) and eigenvalues (the dominance of the principal directions). An eigenvector's large eigenvalue indicates that the data are widely spread along the direction of the eigenvector. PCA reduces the data dimension by selecting eigenvectors corresponding to the dominant eigenvalues. LDFA computes the eigenvectors and eigenvalues of a covariance matrix using singular value decomposition. The number of chosen eigenvectors, N_l, is determined by an energy criterion as shown in Eqs. (21)-(22).

$$E_{N_{in}} = \frac{\sum_{i=1}^{N_{in}} \lambda_{i,in}}{\sum_{i=1}^{\min(N_e, N_\tau)-1} \lambda_{i,\,in}} \quad (21)$$

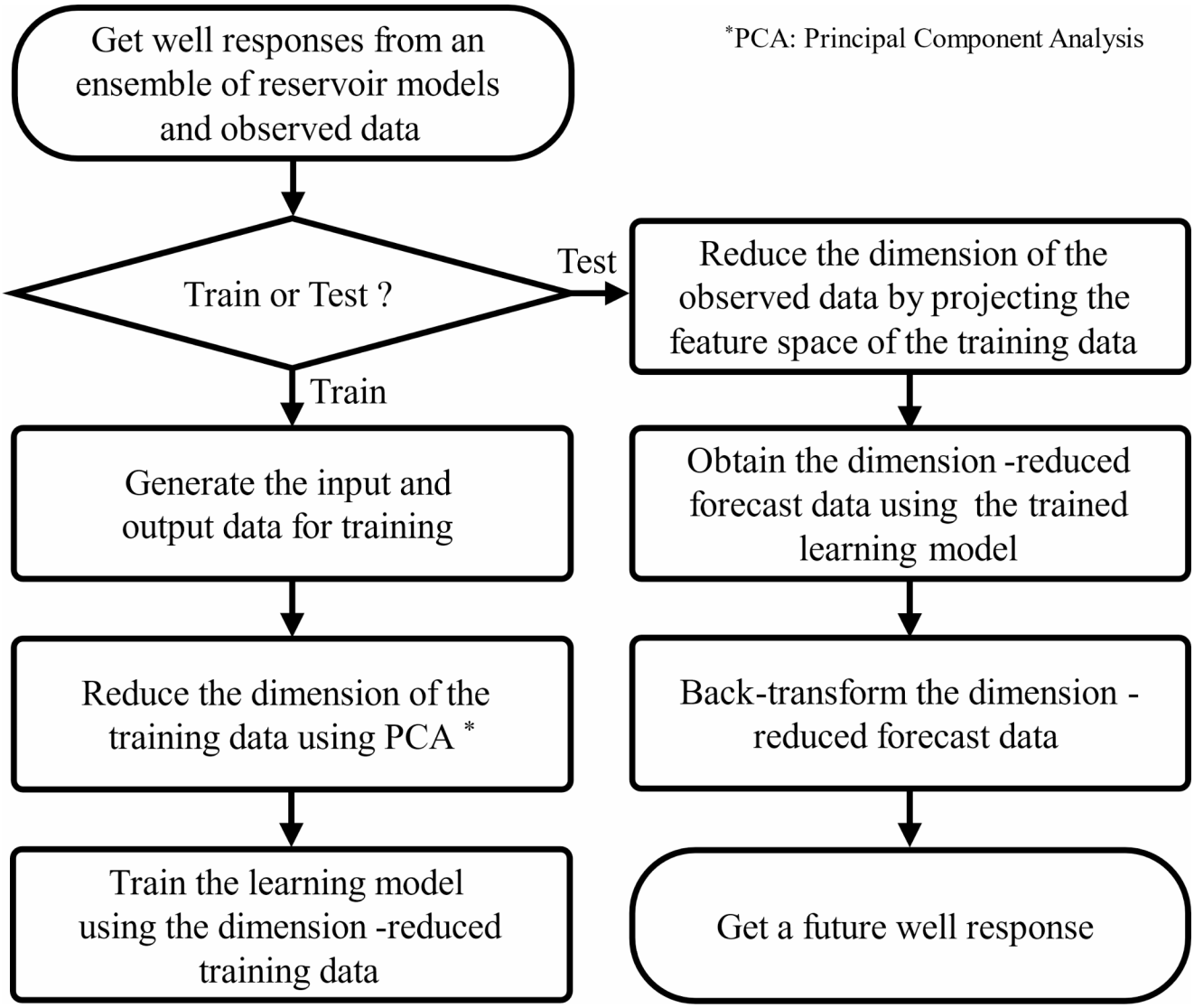


Fig. 2 Flowchart of learning-based data-driven forecast approach (LDFA)

$$E_{N_{out}} = \frac{\sum_{i=1}^{N_{out}} \lambda_{i,\,out}}{\sum_{i=1}^{\min(N_e,N_t)-1} \lambda_{i,\,out}} \tag{22}$$

where $\lambda_{i,\,in}$ and $\lambda_{i,\,out}$ are the *i*-th largest eigenvalues of the training input and output data, respectively, and N_{in} and N_{out} are the numbers of chosen eigenvectors. After reducing the dimensions, LDFA builds a learning model and trains it with the dimension-reduced training input and output data. Then, the future well response is predicted using the observed data. The observed data should be projected onto the dimension-reduced space of the training input data as Eq. (23).

$$x_{obs} = f_1\left(d_{obs}\right), \tag{23}$$

where x_{obs} is the vector of the observations projected on a dimension-reduced space of the training input data, and f_1 is the reparameterization step.

Then, the learning model forecasts future well responses corresponding to the observations as Eq. (24).

$$\widehat{x}_f = \mathcal{L}\left(x_{obs}\right), \tag{24}$$

where $\mathcal{L}$ is the learning model, and $\widehat{x}_f$ is the output vector projected on the dimension-reduced space of the training output data. The prediction results are obtained from the output vector $\widehat{x}_f$ using the back transformation. Then, the prediction results can be expressed using $\widehat{x}_f$ as shown in Eq. (25).

$$\widehat{d}_f = f_1^{-1}\left(\widehat{x}_f\right) \tag{25}$$

where $\widehat{d}_f$ is the prediction results, and f_1^{-1} is a back transform function of f_1.

Multiple forecast processes are conducted to quantify the uncertainty in prediction results. The observed data for the multiple forecast processes are generated as Eq. (26),

$$d_{obs,\,j} = d_{obs} + \epsilon_j,\ j = 1,2,\ \ldots\ ,\ N_f \tag{26}$$

where ϵ_j is a vector of the measurement and modeling errors for the j-th forecast process, and N_f is the number of forecast processes. ϵ_j is sampled from a Gaussian distribution with zero mean and a covariance matrix C_e.

- Learning-based Pattern-driven Forecast Approach.

DSI and LDFA do not provide satisfactory forecast results for the observed data that are outside the range of prior data due to the extrapolation issue. To mitigate the extrapolation issue in DSI and LDFA, a new scaling method is proposed. Figure 3 shows the procedure of LPFA. Compared to LDFA, LPFA adds steps for the new scaling method (dashed line in

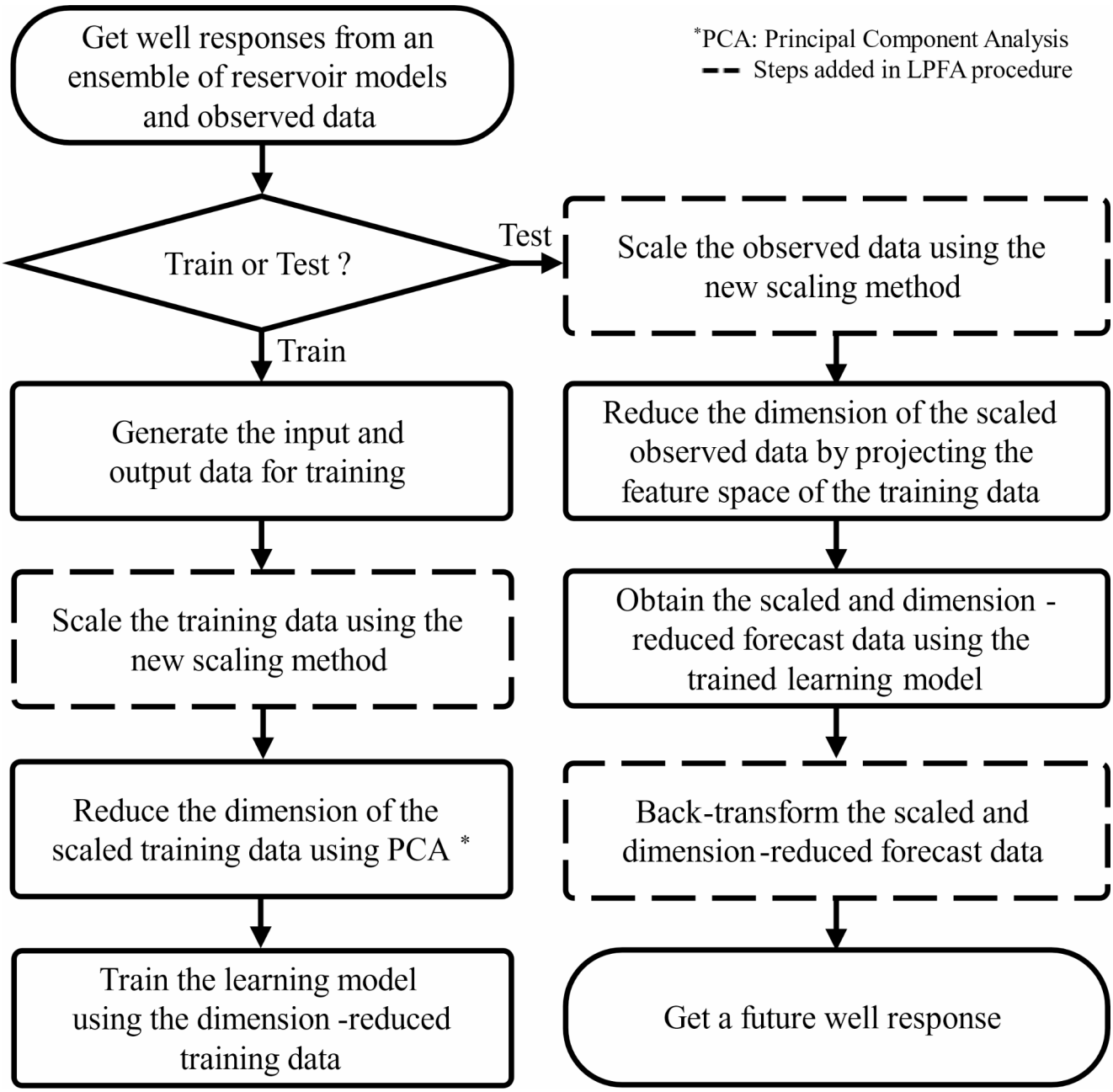


Fig. 3 Flowchart of learning-based pattern-driven forecast approach (LPFA)

Fig. 3). The new scaling method is employed to pattern the prior data before the dimension-reduction step.

The standardization commonly used in ML is typically applied to each feature or variable of a dataset individually, addressing variations in their ranges and distributions that can impact their importance during the learning process. Standardization ensures that all features or variables have comparable ranges and distributions. The training input and output data are scaled by standardization as shown in Eqs. (27)-(32). The standardized training input data, $d_{in,\,standardized}$, and the standardized training output data, $d_{out,\,standardized}$, are expressed as Eqs. (27)-(32).

$$d_{in,\,standardized} = \begin{bmatrix} \frac{d_{\tau_1,1} - \bar{d}_{\tau_1}}{\sigma_{\tau_1}} & \cdots & \frac{d_{\tau_1,N_e} - \bar{d}_{\tau_1}}{\sigma_{\tau_1}} \\ \vdots & \ddots & \vdots \\ \frac{d_{\tau_{N_\tau},1} - \bar{d}_{\tau_{N_\tau}}}{\sigma_{\tau_{N_\tau}}} & \cdots & \frac{d_{\tau_{N_\tau},N_e} - \bar{d}_{\tau_{N_\tau}}}{\sigma_{\tau_{N_\tau}}} \end{bmatrix} \tag{27}$$

$$d_{out,\,standardized} = \begin{bmatrix} \frac{d_{t_1,1} - \bar{d}_{t_1}}{\sigma_{t_1}} & \cdots & \frac{d_{t_1,N_e} - \bar{d}_{t_1}}{\sigma_{t_1}} \\ \vdots & \ddots & \vdots \\ \frac{d_{t_{N_t},1} - \bar{d}_{t_{N_t}}}{\sigma_{t_{N_t}}} & \cdots & \frac{d_{t_{N_t},N_e} - \bar{d}_{t_{N_t}}}{\sigma_{t_{N_t}}} \end{bmatrix} \tag{28}$$

$$\bar{d}_{\tau_i} = \frac{1}{N_e} \sum_{l=1}^{N_e} d_{\tau_i,l},\ i = 1,\ \ldots\ ,\ N_\tau \tag{29}$$

$$\bar{d}_{t_i} = \frac{1}{N_e} \sum_{l=1}^{N_e} d_{t_i,l},\ i = 1,\ \ldots\ ,\ N_t \tag{30}$$

$$\sigma_{\tau_i} = \sqrt{\frac{1}{N_e} \sum_{l=1}^{N_e} \left(d_{\tau_i,l} - \bar{d}_{\tau_i} \right)^2},\ i = 1,\ \ldots\ ,\ N_\tau \tag{31}$$

$$\sigma_{t_i} = \sqrt{\frac{1}{N_e} \sum_{l=1}^{N_e} \left(d_{t_i,l} - \bar{d}_{t_i} \right)^2},\ i = 1,\ \ldots\ ,\ N_t \tag{32}$$

where $\bar{d}_{\tau_i}$ and σ_{τ_i} are the mean and standard deviation at an observation time τ_i, respectively, and $\bar{d}_{t_i}$ and σ_{t_i} are the mean and standard deviation at a forecast time t_i, respectively.

However, standardization cannot address the extrapolation issue for two primary reasons. Firstly, standardization performs scaling for each feature. In the context of predicting future well responses, the well response at each time step corresponds to a feature, and the well response from each reservoir model corresponds to an instance. Therefore, standardization scales the well responses for each time step t_k. If scaling is performed at each time step, the differences in magnitude between well responses remain, which cannot solve the extrapolation issue. Secondly, standardization assumes that the test data follows the distribution of the training data, so the mean and standard deviation of the training input and output data are used to scale the test data. If the observed data is outside the prior data range, the test data does not follow the distribution of the training data. Additionally, scaling with incorrect means and standard deviations degrades prediction performance.

To mitigate the extrapolation issue, a new scaling method is proposed. The new scaling method is similar to standardization but differs in the dimension along which the scaling is performed. Unlike standardization, which performs scaling for each feature or variable, the new scaling method performs scaling for each instance. Furthermore, the new scaling method scales the prior data using the mean and standard deviation of the data during the observation times. This approach is more realistic for predicting future well responses because only data during observation period is available. Additionally, the assumption that the distribution of the test data follows the training data is unnecessary because scaling is performed individually for each reservoir model. The scaled training input data, $d_{in,scaled}$, and scaled training output data, $d_{out,\,scaled}$, are expressed as Eqs. (33)-(36),

$$d_{in,\,scaled} = \begin{bmatrix} \frac{d_{\tau_1,1} - \bar{d}_{\tau,1}}{\sigma_{\tau,1}} & \cdots & \frac{d_{\tau_1,N_e} - \bar{d}_{\tau,N_e}}{\sigma_{\tau,N_e}} \\ \vdots & \ddots & \vdots \\ \frac{d_{\tau_{N_\tau},1} - \bar{d}_{\tau,1}}{\sigma_{\tau,1}} & \cdots & \frac{d_{\tau_{N_\tau},N_e} - \bar{d}_{\tau,N_e}}{\sigma_{\tau,N_e}} \end{bmatrix}, \tag{33}$$

$$d_{out,\,scaled} = \begin{bmatrix} \frac{d_{t_1,1} - \bar{d}_{\tau,1}}{\sigma_{\tau,1}} & \cdots & \frac{d_{t_1,N_e} - \bar{d}_{\tau,N_e}}{\sigma_{\tau,N_e}} \\ \vdots & \ddots & \vdots \\ \frac{d_{t_{N_t},1} - \bar{d}_{\tau,1}}{\sigma_{\tau,1}} & \cdots & \frac{d_{t_{N_t},N_e} - \bar{d}_{\tau,N_e}}{\sigma_{\tau,N_e}} \end{bmatrix}, \tag{34}$$

$$\bar{d}_{\tau,l} = \frac{1}{N_\tau} \sum_{i=1}^{N_\tau} d_{\tau_i,l},\ l = 1,\ \ldots\ ,\ N_e \tag{35}$$

$$\sigma_{\tau,l} = \sqrt{\frac{1}{N_\tau} \sum_{i=1}^{N_\tau} \left(d_{\tau_i,l} - \bar{d}_{\tau,l} \right)^2},\ l = 1,\ \ldots\ ,\ N_e \tag{36}$$

where $\bar{d}_{\tau,l}$ and $\sigma_{\tau,l}$ are the mean and standard deviation, respectively, during the observation times in the l-th ensemble model.

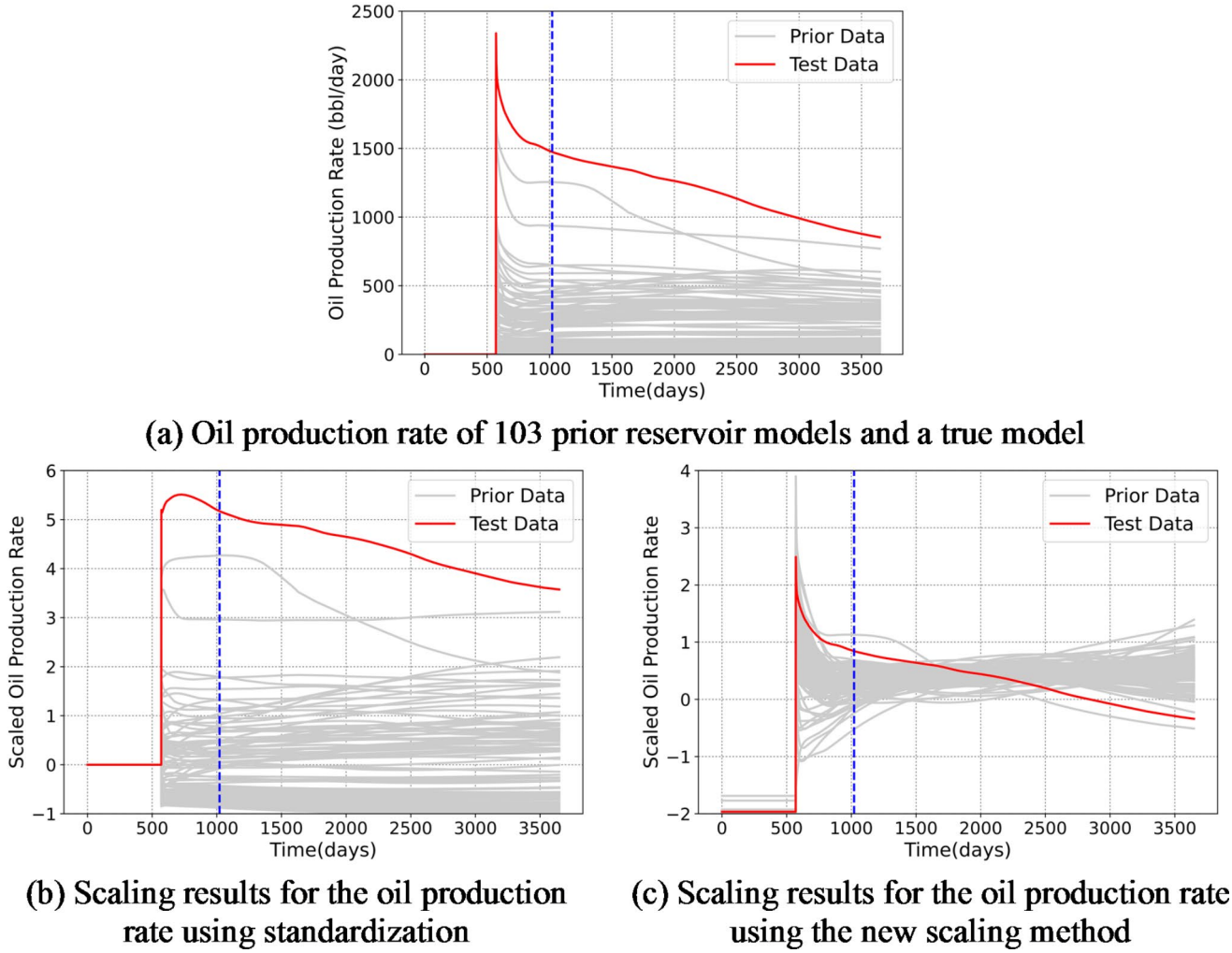
The operational conditions are the same across prior reservoir models for producers and injectors, respectively, resulting in well responses that differ only in magnitude, while patterns often remain similar. Therefore, predicting future well responses based on these patterns can lead to

more effective predictions. Scaling by instance removes the magnitude differences for each instance, revealing the various patterns of well response. This new scaling method enables the learning model to effectively learn patterns, which led to its integration into LDFA and the naming of the approach as the learning-based pattern-driven forecast approach.

Figure 4 shows an example explaining the difference between standardization and the new scaling method. In Fig. 4(a), the red line represents the test data, the oil production rate of a true model, and the gray line represents the prior data, the oil production rate of 103 prior reservoir models. Because the test data is outside the prior data range, DSI and LDFA may provide unsatisfactory prediction results due to the extrapolation issue. Figure 4(b) shows the scaling results using standardization. Standardization scales the oil production rate for each time step, resulting in a similar distribution at each time step after scaling. However, the difference between the test data and the prior data persists, leaving the extrapolation issue unresolved. On the other hand, Fig. 4(c) shows the scaling results using the new scaling method, where the test data is within the range of the prior data. Additionally, the scaled data has a narrow distribution due to the similar pattern of the oil production rate. As a result, the observed data is expected to align closely with the distribution of the prior data after scaling, thereby mitigating the extrapolation issue.

In summary, the new scaling method enables accurate predictions even when observed data is outside the prior data range, without the need to add additional training data. This is because the new scaling method considers the unique characteristics of changes in well responses.

LPFA integrates the new scaling method into LDFA and proceeds as follows. First, forward simulations are conducted on the ensemble of prior reservoir models to obtain prior data, which are then partitioned based on the observation period. Subsequently, each set is scaled using the new scaling method, and the dimension of the scaled data



(a) Oil production rate of 103 prior reservoir models and a true model

(b) Scaling results for the oil production rate using standardization

(c) Scaling results for the oil production rate using the new scaling method

Fig. 4 Example for scaling methods. The gray, red lines, and blue dashed line represent the prior data and test data, and the end of the observation times, respectively

is reduced using PCA. The resulting preprocessed data are used as the training input and output data for the learning model.

- Ensemble Conditioning Step.

The predictions of a single learning model may have high variance due to different initializations and training conditions, leading to inconsistent results. If the learning model is overfitted or converged to a wrong local optimum, it provides unreliable predictions. Methods such as LDFA and LPFA, which rely on a single learning model, are affected by these conditioning and prediction variance issues. To address these issues, an ensemble conditioning step is introduced. The ensemble conditioning step predicts future well responses using multiple learning models, which provide predictions conditioned on the observed data. Figure 5 shows the procedure of the ensemble conditioning step.

ECS consists of two steps: ensemble learning and a conditioning step. Ensemble learning is a technique that uses multiple learning models instead of a single learning model (Dietterich 2000; Sagi and Rokach 2018; Dong et al. 2020).

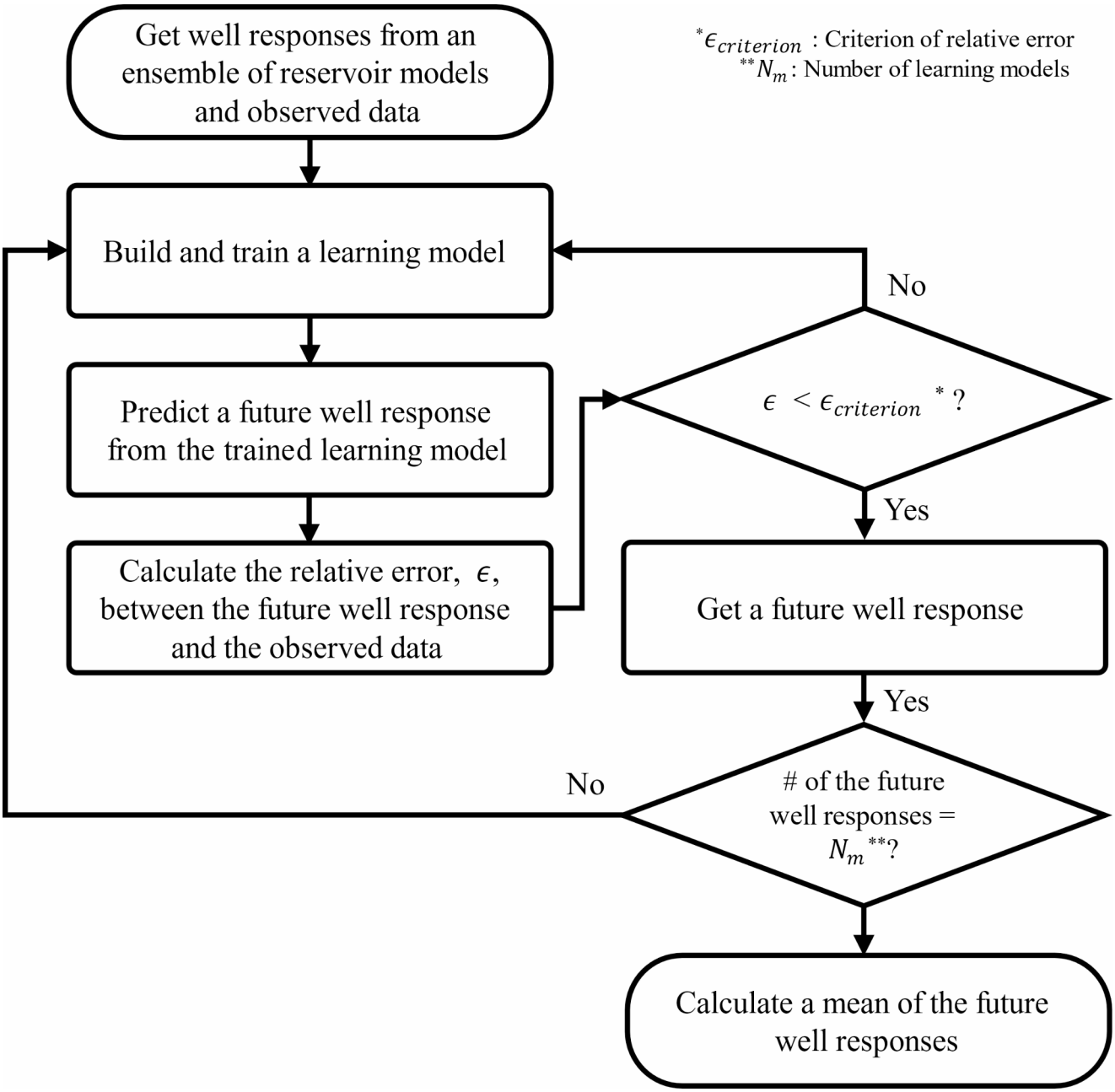


Fig. 5 Flowchart of ensemble conditioning step (ECS)

Using multiple models ensures more accuracy in predicting future well responses compared to using a single learning model. This approach leverages the diversity of the models to improve robustness and reduce the likelihood of overfitting.

In ensemble learning, each model is initialized with different random weights and biases. Random initialization helps in exploring different parts of the loss surface, which can assist in finding better local minima and avoiding suboptimal solutions that might result from poor initialization. Additionally, training and validation data are randomly sampled from prior data. This random sampling ensures that each model in the ensemble learns from a different subset of the data, further enhancing the diversity of the models. As a result, the learning models are trained with different initial weights and biases and different training datasets, effectively implementing ensemble learning. Each model predicts future well responses from the observed data as shown in Eqs. (37)-(38).

$$\widehat{x}_{f,\,k} = \mathcal{L}_k\left(f_1\left(d_{obs,\,true}\right)\right),\ k = 1,\ 2,\ \ldots,\ N_m, \tag{37}$$

$$\widehat{d}_{f,k} = f_1^{-1}\left(\widehat{x}_{f,\,k}\right),\ k = 1,\ 2,\ \ldots,\ N_m, \tag{38}$$

where $\widehat{x}_{f,\,k}$ is the output vector from the k-th learning model, $\widehat{d}_{f,\,k}$ is the prediction result from the k-th learning model, N_m is the number of learning models, and f_1^{-1} is the back transform function of reparameterization. Then, future well responses are the means of the prediction results of the multiple learning models as shown in Eq. (39). Using more learning models provides more consistent prediction results but incurs more computational cost.

$$\widehat{d}_f = \frac{1}{N_m}\sum\nolimits_{k=1}^{N_m}\widehat{d}_{f,\,k} \tag{39}$$

Additionally, a conditioning step is implemented to improve prediction performance. The conditioning step screens predictions that do not honor the observed data sufficiently based on the relative error. The relative error between the prediction and the observed data is calculated using Eq. (40).

$$\epsilon_k = \frac{1}{N_\tau}\sum\nolimits_{i=1}^{N_\tau}\frac{|\widehat{d}_{\tau\,i,f,k} - d_{\tau\,i,\,obs}|}{d_{\tau\,i,\,obs}}, \tag{40}$$

where ϵ_k is the relative error of the k-th learning model, $\widehat{d}_{\tau\,i,\,f,\,k}$ is the prediction results of the k-th learning model at τ_i, and $d_{\tau\,i,\,obs}$ is the observations at τ_i. If the relative error exceeds the criterion, $\epsilon_{criterion}$, predictions with $\epsilon_k > \epsilon_{criterion}$ are discarded.

Results

- Brugge Model.

The prediction performances of DSI, LDFA, LPFA, and ECS are compared using the Brugge model. The Brugge model is a benchmark model for testing different waterflooding optimization and history-matching methods (Peters et al. 2010). The Brugge Field, a synthetic oil field, is characterized by an elongated east-west-oriented half-dome structure, which includes a large boundary fault and an internal fault. The size of the Brugge model is about 17,000 m × 9,000 m × 60 m with approximately 60,000 grid blocks. The simulation model comprises nine layers, each containing 139 × 48 grid blocks, totaling 44,550 active cells. All results presented in this paper are derived from this simulation model. Figure 6 shows the top depth of the Brugge model and the well locations. The Brugge model has 10 injectors and 20 producers. In Fig. 6, 'P' denotes a producer, and 'I' denotes an injector. P9, production well numbered 9, is operated at a constant bottom-hole pressure (725.5 psi), and other producers are operated at a constant liquid production rate (about 2,000 STB/day). The injectors are operated at a constant water injection rate (4,000 STB/day).

The Brugge model consists of four formations arranged from top to bottom: Schelde, Maas, Waal, and Schie. There is no continuous shale barrier between different formations. These formations have differences in the distribution of rock facies, porosity, horizontal and vertical permeabilities, net-to-gross, and initial water saturation. Table 1 shows the average reservoir properties and the depositional environment of the four formations.

The Netherlands Organization for Applied Scientific Research (TNO) developed the Brugge model and provided 104 realizations of geological models. All of the realizations have the same fault, structure, and grid system but they have different rock facies, porosity, horizontal and vertical permeabilities, net-to-grosses, and initial water saturations. The 104 realizations were generated using different geostatistical techniques such as sequential Gaussian simulation (SGS), object-based modeling, and sequential indicator simulation (SIS). Figure 7 shows three realizations among the entire 104 ones. Realizations 1, 28, and 66 shown in Fig. 7 were generated using SGS, object-based modeling, and SIS, respectively. The 104 realizations were simulated for ten years using the black oil reservoir simulator, ECIPSE 100 (Schlumberger 2021). These 104 instances, simulation results of the 104 realizations, are used as training and test data.

- Comparison of DSI, LDFA, and LPFA.

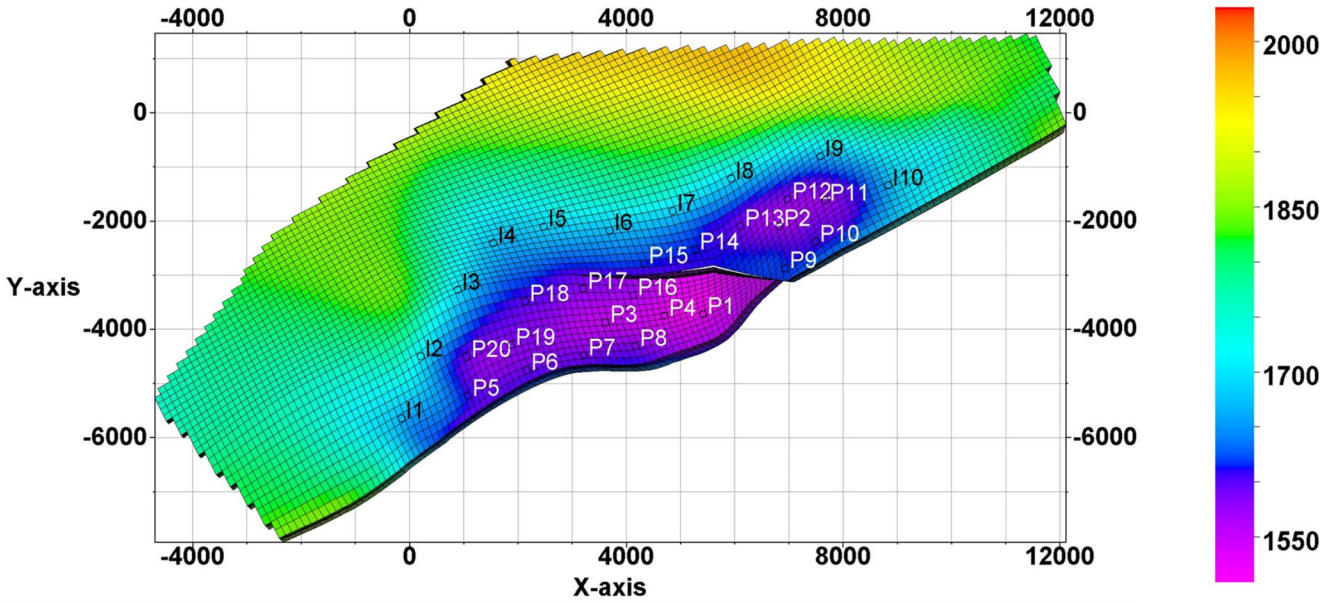


Fig. 6 Top depth (m) of the Brugge model and the well locations. The injectors and producers are labeled in black and white with I and P prefixes, respectively

Table 1 Reservoir properties and depositional environment of the four formations of the Brugge model

Formation	Schelde	Maas	Waal	Schie
Average thickness (m)	10	20	26	5
Average porosity (%)	20.7	19.0	24.1	19.4
Average permeability (md)	1105	90	814	36
Average net-to-gross (%)	60	88	97	77
Depositional environment	Fluvial	Lower shoreface	Upper shoreface	Sandy shelf

To compare the performances of DSI, LDFA, and LPFA, a realization is chosen as the true model, while the others serve as prior models for generating prior data. The test is conducted 104 times, each time selecting a different realization as the true model. For the ML algorithms of LDFA and LPFA, a feed-forward network with one hidden layer was used. The results of these tests are presented in Tables 2 and 3. The simulation results of the true model and prior models are used as the true well response and prior well responses, respectively. Each well response includes oil and water production rates and bottom-hole pressures at 498 dates (t_1, t_2 ,…, t_{498} = 0, 1,…, 3648 days). Among the 498 dates, only 8 dates (τ_1, τ_2,…, τ_8=16, 151, 301, 451, 580, 703, 852, 1004 days) are used as the observation times. As described in Eqs. (21) and (22), the principal directions are chosen using the energy criterion of PCA ($E_{N_{in}}$ and $E_{N_{out}}$= 0.99) to reduce the dimension of the prior data. The uncertainty in the prediction results is quantified by repeating the forecast process 300 times. The observed data used in each forecast process are perturbed by adding Gaussian random noises to the well responses of the true model, as expressed in Eq. (5). The standard deviation of the Gaussian random noise is 2% of the observations. As different observed data are used in each repetition, 300 different prediction results are obtained. Then, the 10th and 90th percentiles (P_{10} and P_{90}) of the prediction results are used as the uncertainty interval in the results.

The prediction performances of DSI, LDFA, and LPFA are evaluated under two scenarios: (i) prior data include well responses close to the observed data, and (ii) prior data do not include well responses close to the observed data. To visualize whether the prior data include well responses close to the observed data, the prior data and the true response are visualized in a two-dimensional feature space obtained by principal component analysis (PCA). In the following figures, a red triangle represents the true well response and gray circles represent the well responses of the prior models. In the feature space, the x- and y-axes represent the first and second principal directions of the prior data with the largest eigenvalues, respectively. In the following results, the eigenvalues corresponding to first and second principal directions are significantly larger than the other eigenvalues. Therefore, the original data are sufficiently described using two principal directions. In a two-dimensional feature space, similar data are located close to each other.

Figure 8(a) and (b) show the feature space of the bottom-hole pressure at P5 and the oil production rate at P9, respectively. As the true response in Fig. 8(a) and (b) are located within the prior data, these scenarios correspond to case (i).

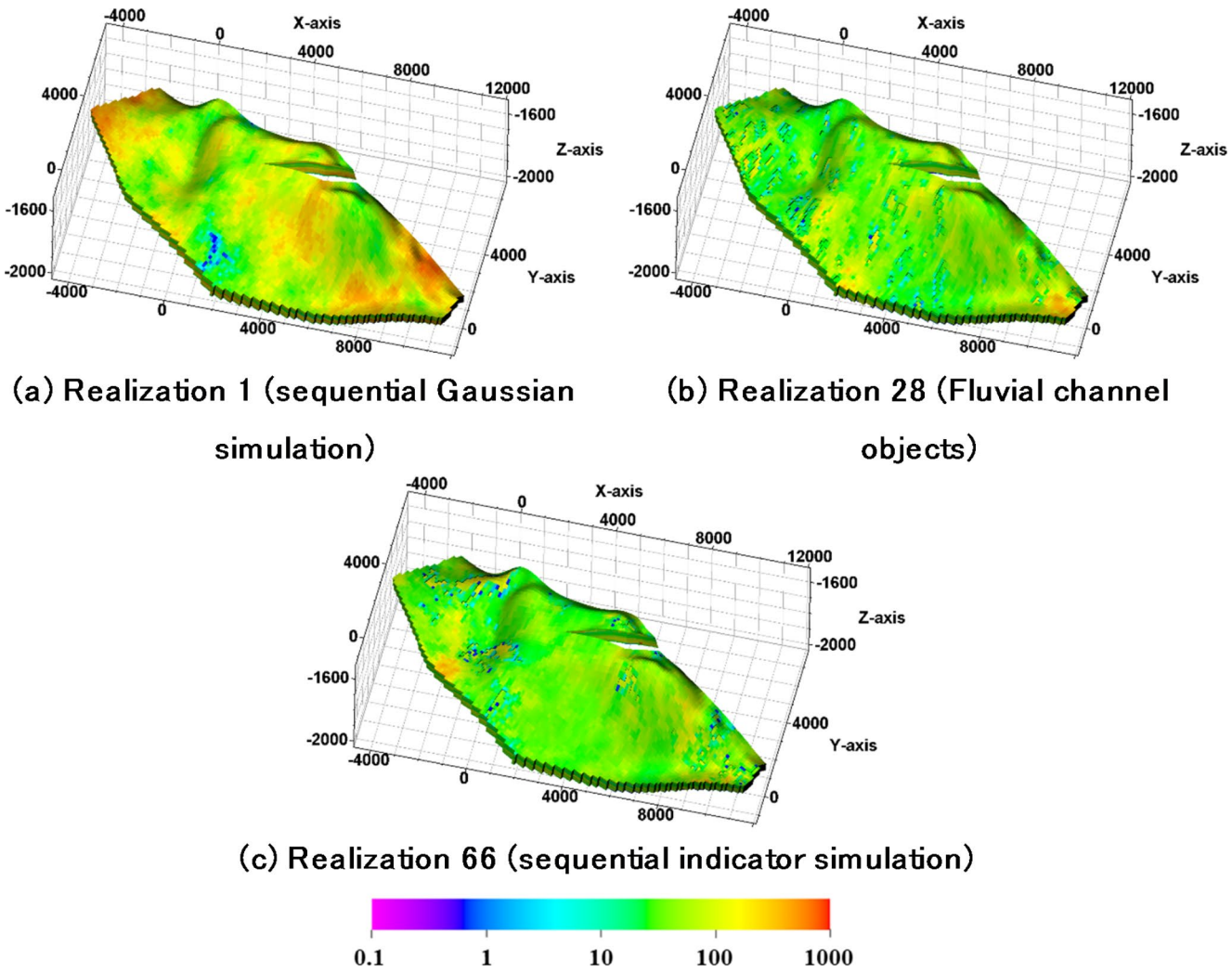


Fig. 7 Three realizations of the log (permeability (md)) in the x-direction. (a) Realization 1, (b) Realization 28, and (c) Realization 66 are generated using SGS, object-based modeling, and SIS, respectively

Table 2 Median relative errors of DSI, LDFA, and LPFA. WBHP, WOPR, and WWPR represent bottom-hole pressure, oil production rate, and water production rate, respectively

Well name and response	DSI (%)	LDFA (%)	LPFA (%)
P1 WBHP	1.14	2.16	1.08
P2 WBHP	1.20	0.29	0.48
...	…	…	…
P20 WBHP	1.04	0.93	1.42
P5 WOPR	2.83	3.98	2.36
P9 WOPR	6.18	12.23	3.56
...	…	…	…
P20 WOPR	3.41	2.63	2.03
P1 WWPR	4.03	11.71	2.65
P5 WWPR	2.73	6.70	2.10
...	…	…	…
P20 WWPR	1.41	0.80	3.25
Mean of median relative error	2.49	2.85	2.28

Table 3 Mean relative errors of DSI, LDFA, and LPFA.WBHP, WOPR, and WWPR represent bottom-hole pressure, oil production rate, and water production rate, respectively

Well name and response	DSI (%)	LDFA (%)	LPFA (%)
P1 WBHP	6.65	7.90	2.96
P2 WBHP	4.01	2.38	0.78
...	…	…	…
P20 WBHP	3.93	1.33	1.44
P5 WOPR	6.37	7.73	3.67
P9 WOPR	18.20	39.28	5.05
...	…	…	…
P20 WOPR	6.98	6.67	2.18
P1 WWPR	9.81	23.36	3.39
P5 WWPR	7.32	16.70	3.25
...	…	…	…
P20 WWPR	4.25	4.19	4.36
Mean of mean relative error	9.28	11.39	4.89

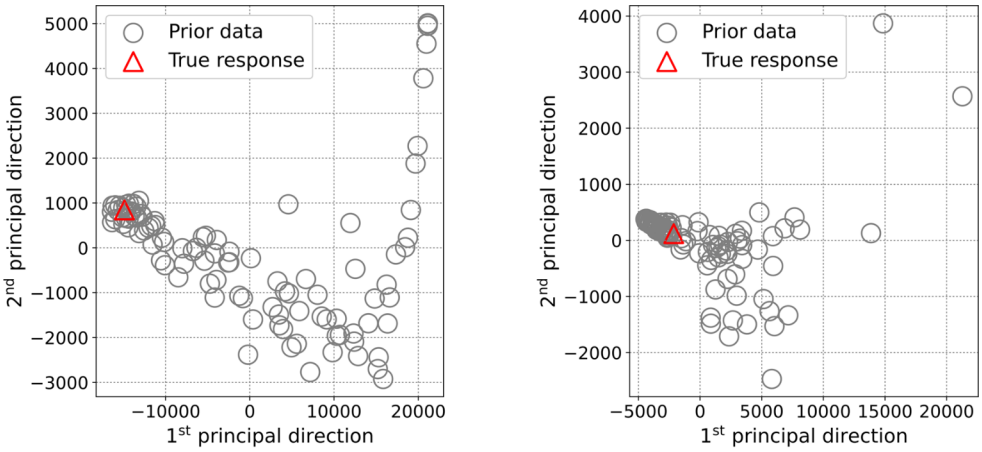


Fig. 8 Feature spaces of the true response and the prior data that include well responses close to the observed data. The x- and y-axes represent first and second principal directions of the prior data with the largest eigenvalues

Figures 9 and 10 present the P_{10} and P_{90} of the prediction results made by DSI, LDFA, and LPFA for the cases shown in Fig. 8(a) and (b), respectively. In the following figures, a red line (True) represents the true well response, gray lines represent the well responses of the prior models, a red diamond (Obs) represents the observed data, a blue dashed line represents the end of the observation times, and black dashed lines (Estimates) represent the P_{10} and P_{90} of the predictions. For the prediction to be reliable and accurate, black dashed lines should honor the red line with acceptable uncertainty range. As shown in Figs. 9 and 10, the true well response lies within the uncertainty intervals of the prediction results. Because the prior data include well responses close to the observed data, DSI, LDFA, and LPFA were able to accurately predict the future well response.

Figure 11(a) and (b) show the feature space of the bottom-hole pressure at P6 and the oil production rate at P5, respectively. (Figure 12) As the true well response falls outside the prior data, these scenarios correspond to case (ii). Figures 13 and 14 show the P_{10} and P_{90} of the prediction results made by DSI, LDFA, and LPFA for the cases depicted in Fig. 11(a) and (b), respectively.

In case (ii), as shown in Figs. 13 and 14, reliable predictions are not achieved by DSI and LDFA due to the extrapolation issue. In contrast, accurate predictions are provided by LPFA, even when the prior data do not include well responses close to the observed data. This improvement is achieved by the implementation of the new scaling step. Figure 12 shows the feature space of the scaled true response and the scaled prior data. The new scaling step removes the magnitude differences for each instance, thereby bringing the true response within the range of the prior data.

Additionally, the performances of DSI, LDFA, and LPFA are compared across the entire ensemble models of the Brugge models. The forecast performance is calculated using the two distinct metrics. The first metric is the median relative error, which is the median of the average of the relative errors over all times. The second metric is the mean relative error, which is the median of the average of the relative errors over all times. The median relative error and mean relative error are defined as Eq. (41) and Eq. (42).

$$Median\ relatvie\ error = median_{j=1,\ldots,N_e}\left(\frac{1}{t}\sum_{i=1}^{N_t}\frac{|d_{t_i,j}-d_{t_i,true}|}{d_{t_i,true}}\right), \tag{41}$$

$$Mean\ relatvie\ error = mean_{j=1,\ldots,N_e}\left(\frac{1}{t}\sum_{i=1}^{N_t}\frac{|d_{t_i,j}-d_{t_i,true}|}{d_{t_i,true}}\right), \tag{42}$$

The median relative error provides an overview of the typical prediction accuracy of each method. Because the median is not sensitive to outliers, it effectively reflects the general performance of the methods across most cases. Table 2 shows the median relative errors of DSI, LDFA, and LPFA. Considering the noise level (2%) in the observed data, the accuracies of DSI, LDFA, and LPFA with an average of 2∼3% of median relative error are acceptable.

The mean relative error, on the other hand, is sensitive to outliers. It reflects the performance of each method, including cases where the observed data fall outside the range of prior data. Table 3 shows the mean relative errors of DSI, LDFA, and LPFA. Considering the case of predicting well responses outside the range of prior data, LPFA maintains a prediction error of 4.89%, unlike other methods whose performance significantly declines.

In summary, while DSI and LDFA provide satisfactory performance in most scenarios, their performance degrades when observed data is outside the range of prior data. In contrast, LPFA maintains its performance across different conditions, making it a more robust choice for predictions in scenarios where observed data may not always be within the range of prior data.

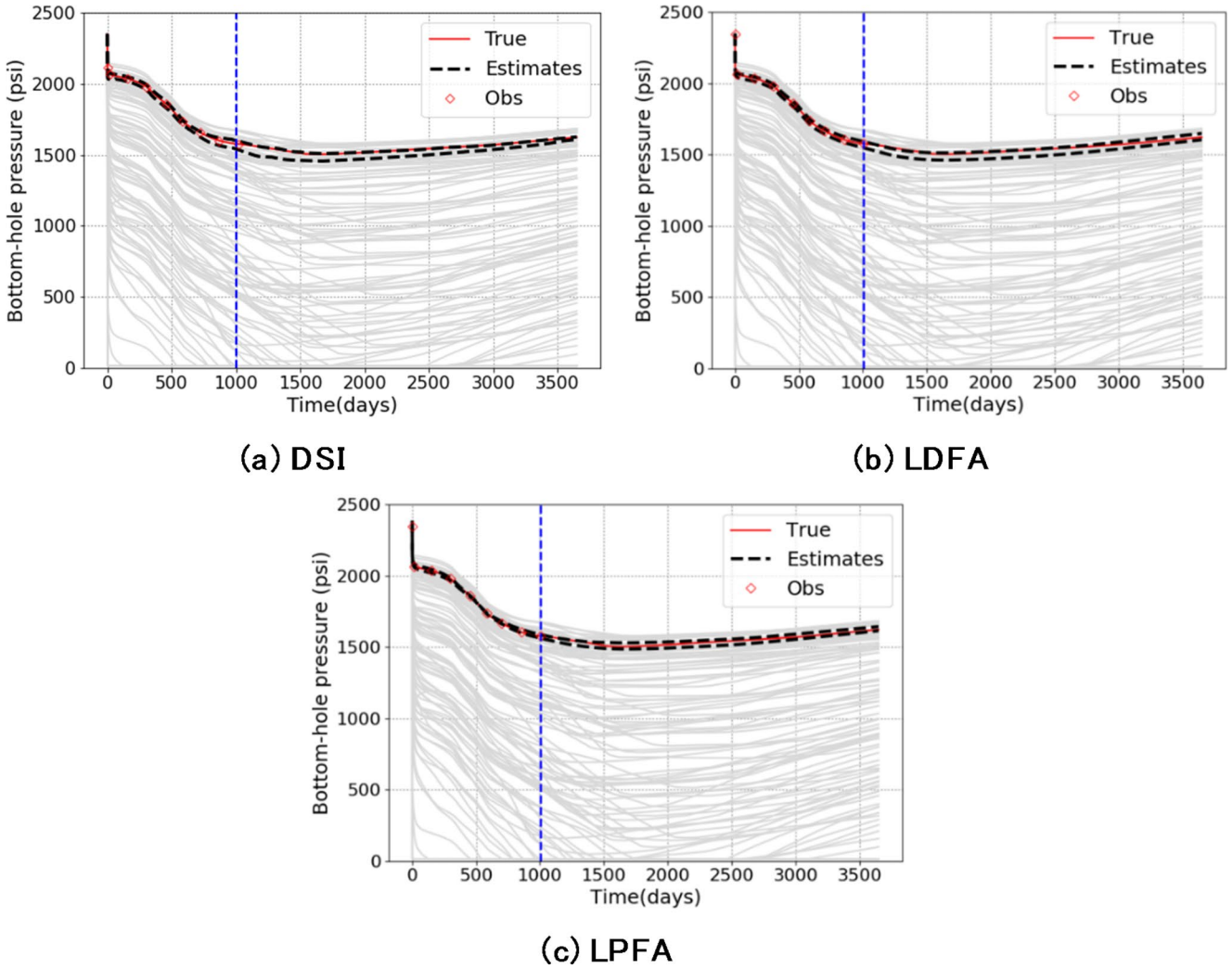


Fig. 9 Prediction results of P5 bottom-hole pressure using (a) DSI, (b) LDFA, and (c) LPFA. Gray lines represent the well responses of prior models, and blue dashed line marks the end of the observation times

- Ensemble Conditioning Step.

The performances of LDFA and LPFA, both with and without ECS, are compared. For ensemble learning, ten learning models (N_m=10) are used, as expressed in Eq. (37). In the conditioning step, a relative error criterion of 4% ($\epsilon_{criterion} = 0.04$) is applied, as described in Eq. (40). To verify that ECS improves the prediction performances of LDFA and LPFA, the prediction process is repeated four times, both with and without ECS. Since the learnable variables of each learning model, such as weights and biases, differ across the four repetitions, the learning models yield different predictions.

Figures 15 and 16 show the prediction results of LDFA without ECS and with ECS, respectively. In Fig. 15, the true well response does not consistently lie within the uncertainty interval of the prediction results. This inconsistency arises because each learning model starts with different initial weights and biases, leading to variations in the optimal weights and biases obtained during training. In Fig. 15 (a), (c), and (d), the predictions fail to sufficiently honor the observed data. While the predictions in Fig. 15 (b) align well with the observations, the appropriate learning model honoring the observations is not always guaranteed.

In Fig. 16, the true well response lies within the uncertainty interval of the prediction results for all predictions. This is attributed to ECS applied to LDFA, which ensures predictions honor the observations in multiple learning models. Consequently, ECS mitigates the conditioning and variance issues of LDFA predictions. Similarly, ECS enhances the prediction performance of LPFA, as illustrated in Figs. 17 and 18.

To quantitatively compare performance, the improvement in prediction accuracy of LPFA with and without ECS is evaluated using the oil production rate data. The prediction accuracy improves from 4.49 to 3.27% due to the

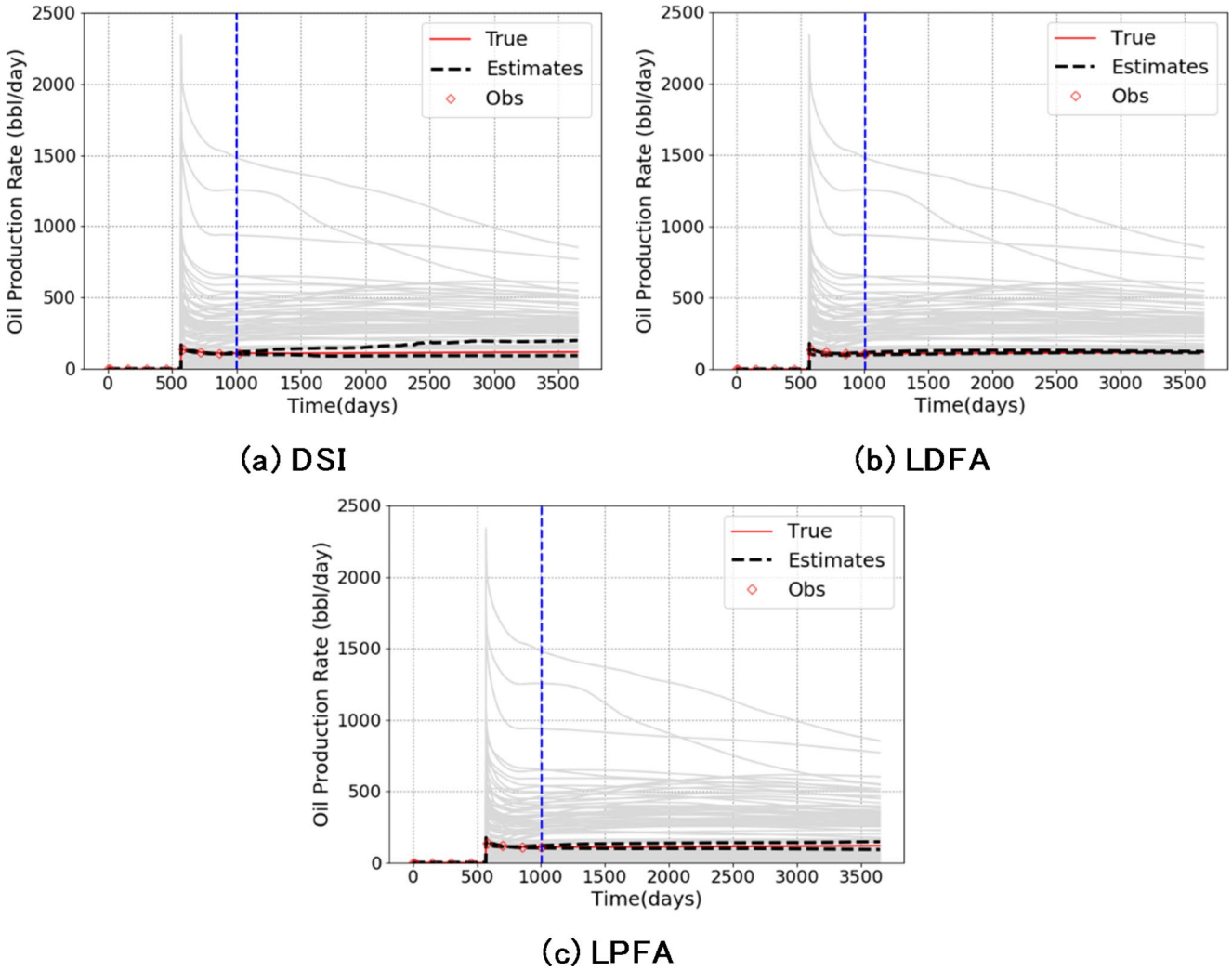
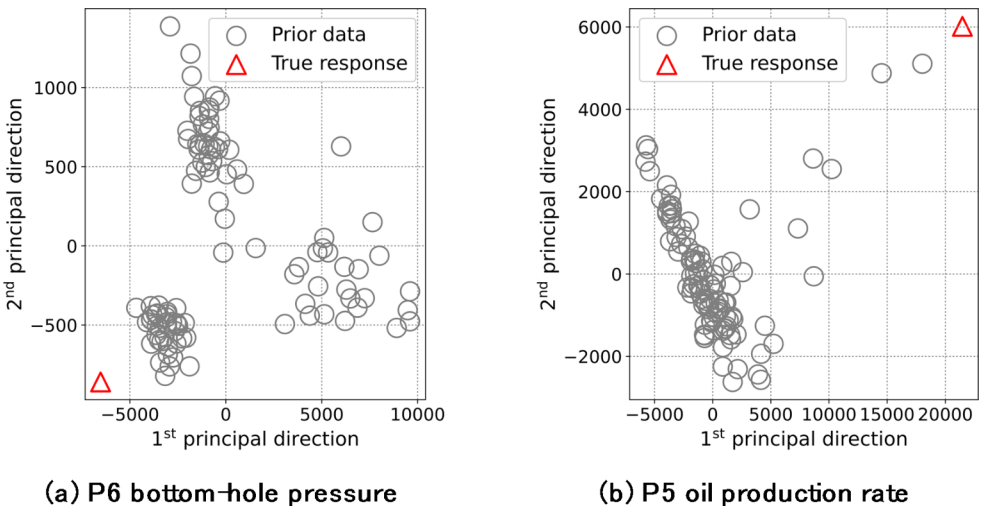


Fig. 10 Prediction results of P9 oil production rate using (a) DSI, (b) LDFA, and (c) LPFA. Gray lines represent the well responses of prior models, and blue dashed line marks the end of the observation times

Fig. 11 Feature spaces of the true response and the prior data that do not include well responses close to the observed data. The x- and y-axes represent first and second principal directions of the prior data with the largest eigenvalues



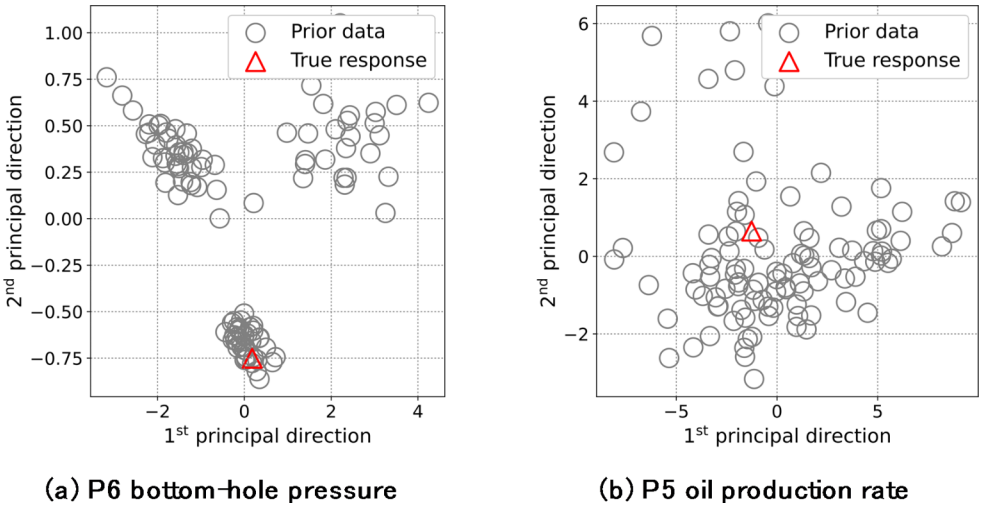


Fig. 12 Feature spaces of the scaled true response and the scaled prior data that include well responses close to the observed data. The x- and y-axes represent first and second principal directions of the scaled prior data with the largest eigenvalues

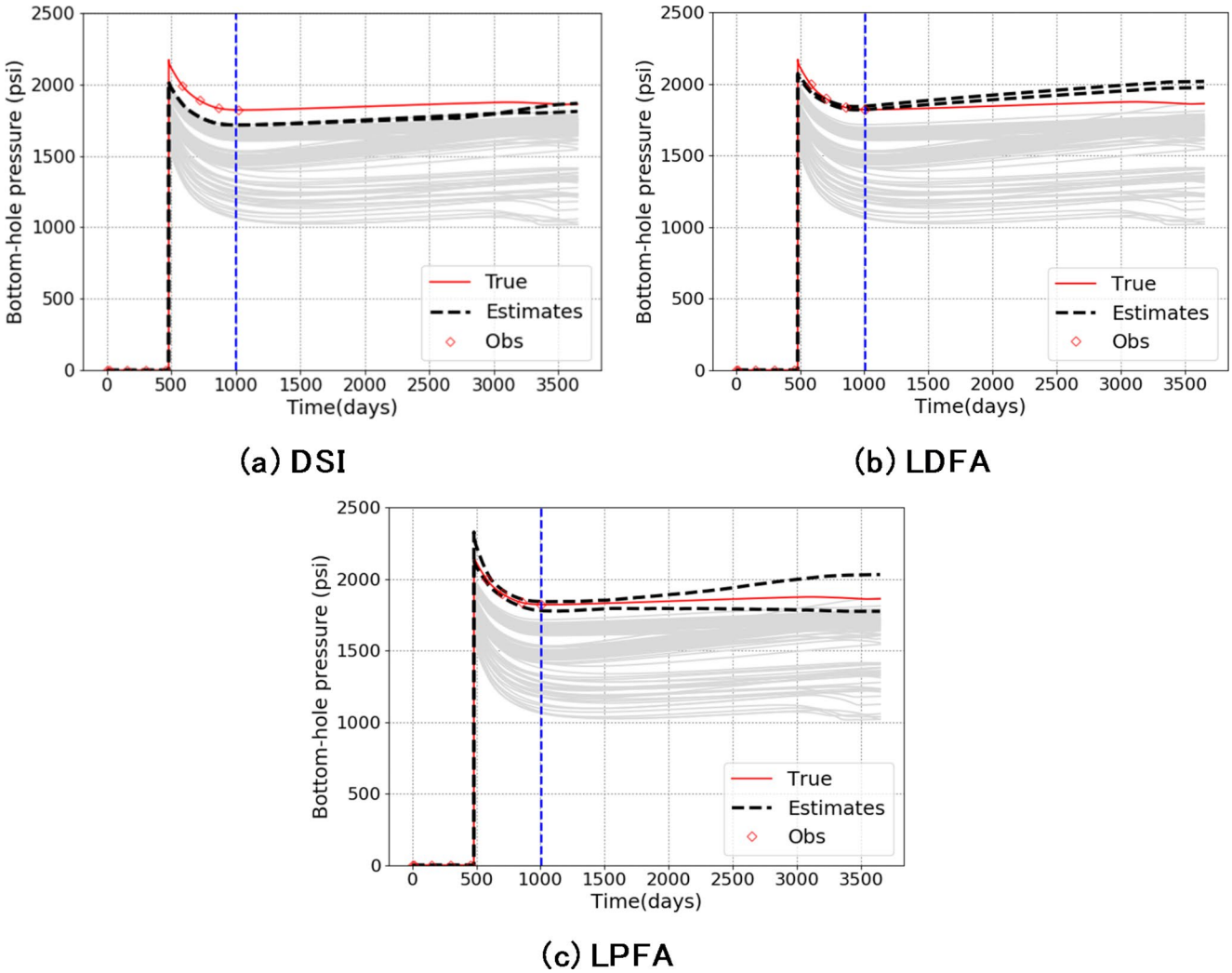


Fig. 13 Prediction results of P6 bottom-hole pressure obtained using DSI, LDFA, and LPFA. Gray lines represent the well responses of prior models, and blue dashed line marks the end of the observation times

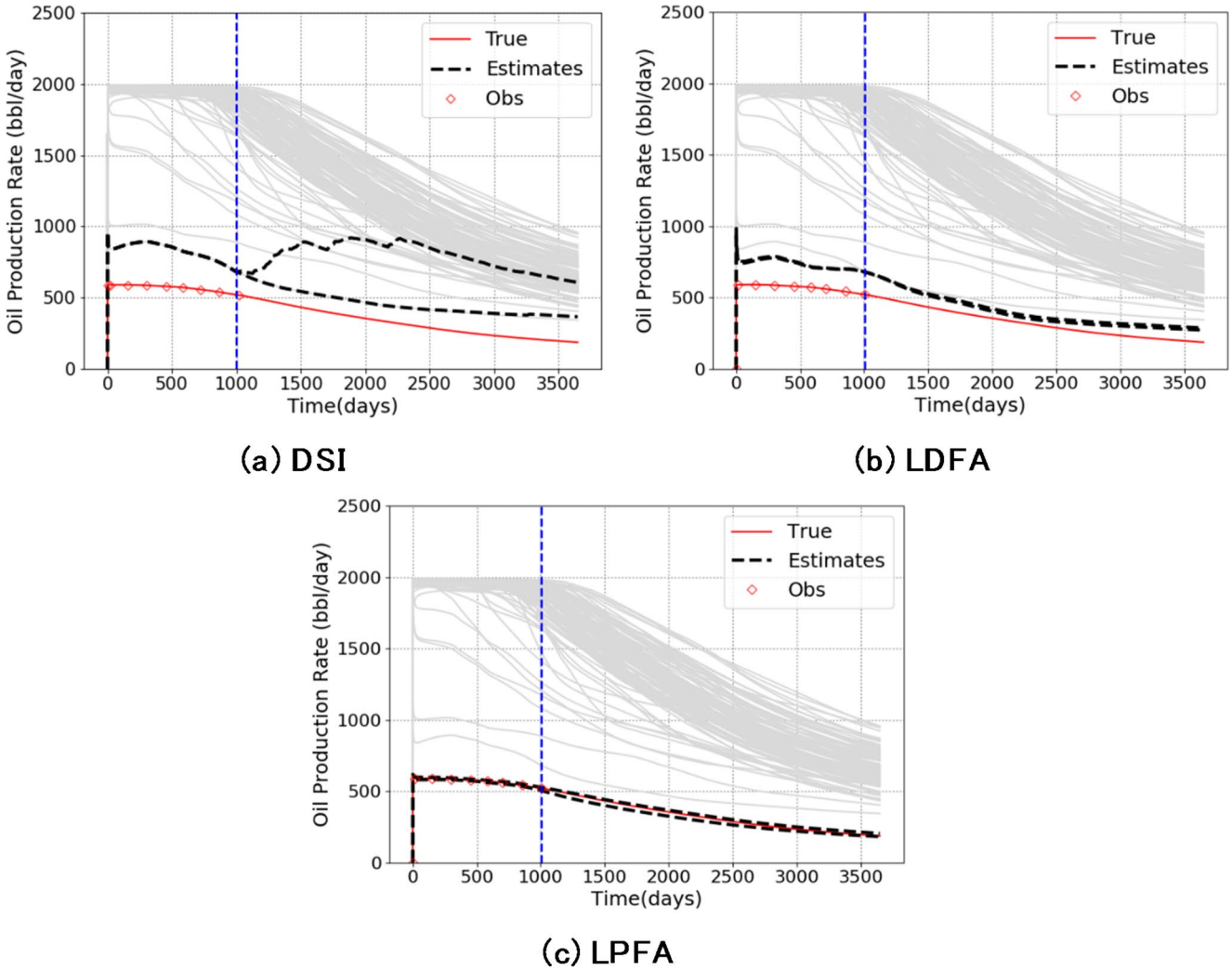


Fig. 14 Prediction results of P5 oil production rate obtained using DSI, LDFA, and LPFA. Gray lines represent the well responses of prior models, and blue dashed line marks the end of the observation times

conditioning step. Furthermore, the uncertainty in future well response predictions is reduced by approximately 50% as a result of ensemble learning.

- Sensitivity Analysis.

In this section, a sensitivity analysis is conducted to evaluate the influence of various factors on the prediction results of the proposed method (LPFA with ECS). The sensitivity analysis involves predicting future well responses under varying observation errors, observation periods, prediction periods, and production strategies.

In the base case, the observation error is set to a noise level of 2% (ϵ = 2%), the observation period is up to 1,000 days, the prediction period is 10 years, and the production strategy is a constant liquid rate. Figure 19(a) presents the uncertainty interval of the prediction results for the oil production rate at P10 under these base case settings.

Figure 19(b) shows the effect of a higher noise level (ϵ = 4%) on the reliability of the uncertainty interval. Compared to the base case (ϵ = 2%), the uncertainty interval widens because the perturbed observations deviate further from the true data, increasing the uncertainty in the input data.

Figure 19(c) examines the impact of extending the observation period (to 1,500 days) on the uncertainty interval. The results show a narrower uncertainty interval compared to the base case, as the additional observed data enhances prediction reliability by reducing the uncertainty in the input data.

Figure 19(d) evaluates the effectiveness of the proposed method for long-term prediction (over 30 years). To demonstrate this, a flow simulation is conducted for an additional 20 years under the same production and injection strategy. Predictions are made using observations up to 1,000 days, consistent with the base case. The results confirm

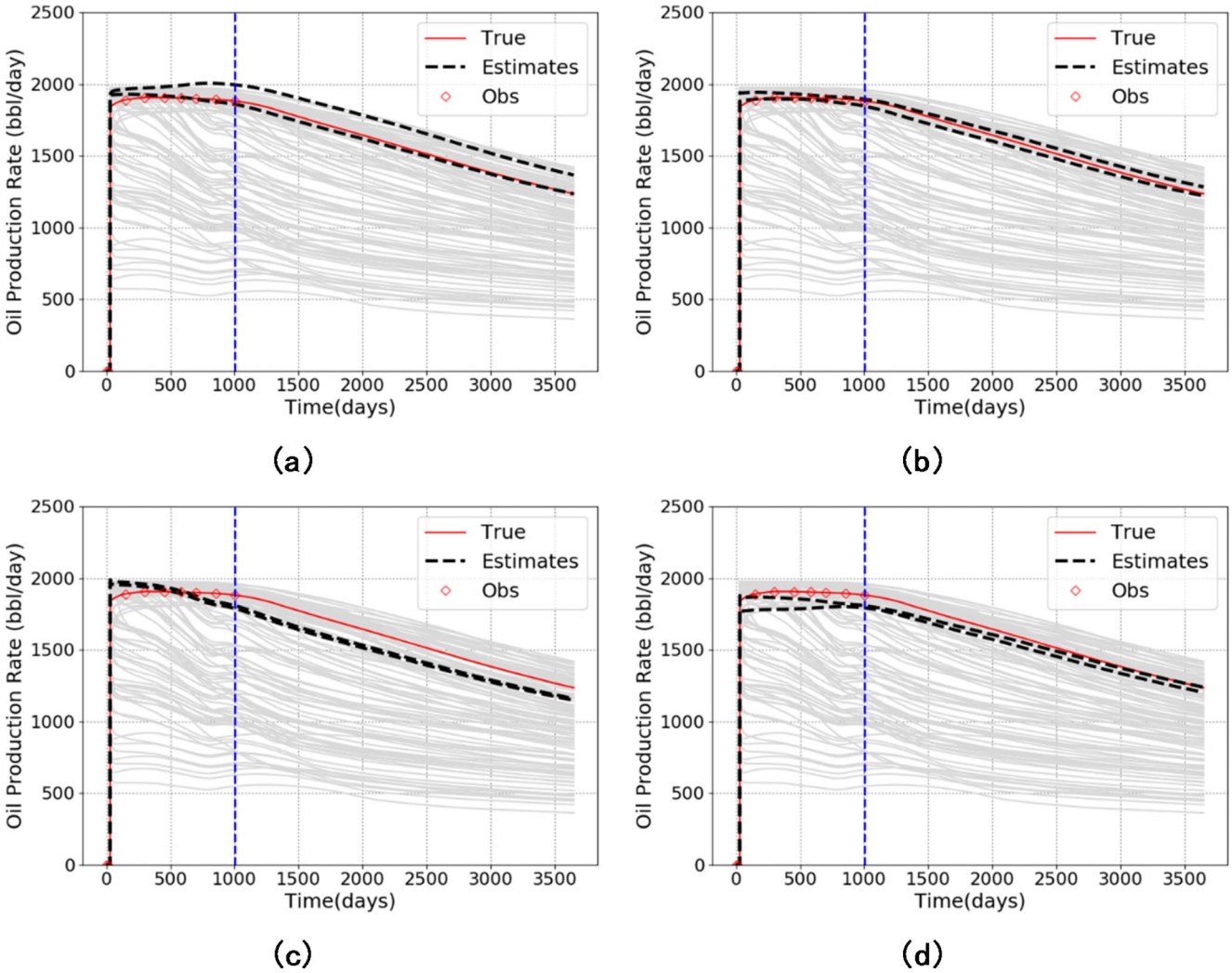


Fig. 15 Four predictions of P10 oil production rate using LDFA without ECS. Gray lines represent the well responses of prior models, and blue dashed line marks the end of the observation times

that the proposed method remains effective for long-term predictions.

Figure 19(e) explores the impact of changing the production strategy on the prediction performance. The production strategy of P10 is altered from a constant liquid rate (approximately 2,000 STB/day) to a constant bottom-hole pressure (1,000 psi). The results underscore the adaptability of the proposed method to changes in production strategies.

- Olympus Data Set.

The prediction performances of DSI, LDFA, and LPFA are evaluated using the Olympus model, which serves as a benchmark for field development optimization under uncertainty (Fonseca et al. 2020). The synthetic reservoir model, Olympus, is inspired by Brent-type oil fields in the North Sea. The field spans approximately 9,000 m x 3,000 m and includes a large boundary fault as well as six internal faults. The reservoir, which is 50 m thick, is divided into two zones by an impermeable shale layer: the upper zone contains fluvial channel sands, while the lower zone consists of clinoforms composed of alternating layers of coarse, medium, and fine sands. The permeability distribution for the upper and lower formations, illustrated in Fig. 20, highlights the presence of highly permeable channels in the upper formation and less heterogeneity in the lower formation.

The model comprises grid cells approximately 50 m x 50 m x 3 m each, totaling 341,728 cells, with 192,750 cells being active. It incorporates five facies types across different layers, with petrophysical properties such as porosity, permeability, and net-to-gross generated using standard geostatistical techniques. Table 4 show the reservoir properties for the different facies.

No porosity-permeability relationship was assumed due to limited early-stage data availability. The ratio between vertical permeability and horizontal permeability is 0.1. The

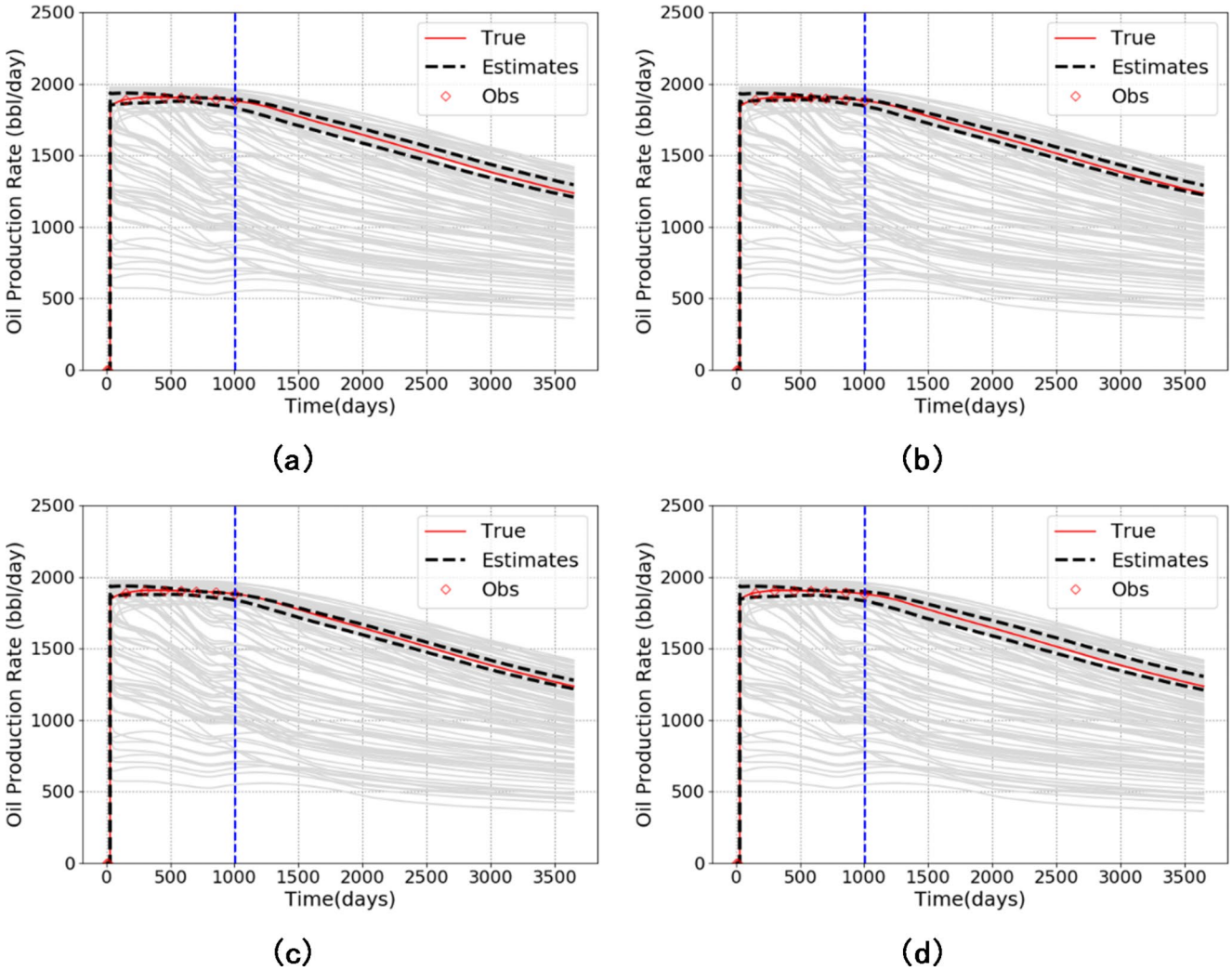


Fig. 16 Four predictions of P10 oil production rate using LDFA with ECS. Gray lines represent the well responses of prior models, and blue dashed line marks the end of the observation times

oil-water contact (OWC) is determined as 2,092 m with an in-situ hydrostatic pressure of 205 bar based on the well logs. The model also considers initial water saturation distribution based on capillary pressure curves and fluid properties such as oil viscosity and formation volume factor.

An ensemble of 50 realizations was generated, varying facies by altering the random seed. Known properties such as grid and fault geometries were consistent across all realizations. Uncertain properties included facies, porosity, permeability, NTG, initial water saturation, and transmissibility across faults. The ensemble was created by generating a high-fidelity base model of approximately 5 million grid cells, drilling five wells, and generating synthetic logs to inform the ensemble generation. Each high-fidelity model was upscaled for flow simulations using a flow-based upscaling method (Gautier et al. 1999; Cao and Kitanidis 1999; Ebrahimi and Sahimi 2004; He and Durlofsky 2006; Hassanpour et al. 2010), maintaining a constant OWC location across realizations. The Olympus model includes 11 producers and 7 injectors operated at constant bottom-hole pressures of 150 and 235 bar, respectively, for 20 years. Figure 21 illustrates the top depth of the Olympus model and the locations of the wells. These 50 realizations are simulated using ECIPSE 100. The simulation results of these 50 realizations serve as training and test data.

To compare the prediction performances of DSI, LDFA, and LPFA, one realization is selected as the true model, and others are used as prior models. The simulation results of the true and prior models are used as the true well and prior well responses, respectively. Each well response includes the oil and water production rates and bottom-hole pressures at 241 dates ($t_1, t_2,\ldots, t_{241} = 0, 31,\ldots, 7304$ days). Among the 241 dates, 12 dates ($\tau_1, \tau_2,\ldots, \tau_{12}$=0, 181, 365, 546, 730, 911, 1095, 1276, 1460, 1642, 1826, 2007 days) are used as observation times. Other parameters for DSI, LDFA, and LPFA are identical to those used in the Brugge model.

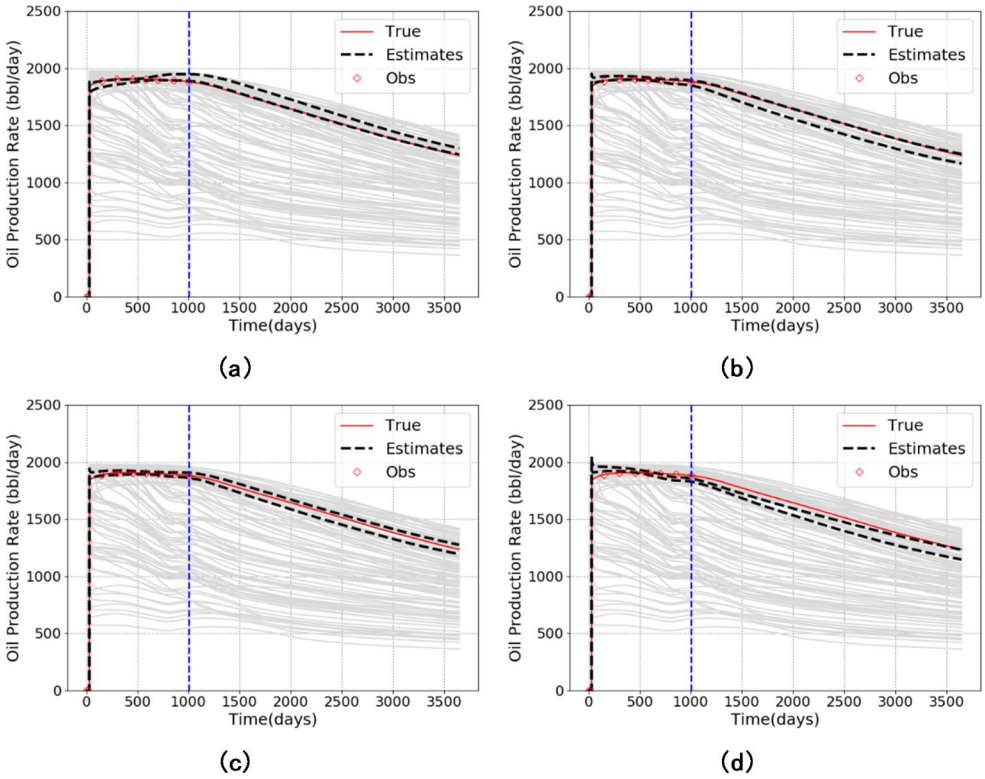


Fig. 17 Four predictions of P10 oil production rate using LPFA without ECS. Gray lines represent the well responses of prior models, and blue dashed line marks the end of the observation times

As in the Brugge model, the prediction performances of DSI, LDFA, and LPFA are compared for two cases: (i) prior data include well responses close to the observed data, and (ii) prior data do not include well responses close to the observed data. Ten learning models (N_m=10) and 4% as the criterion for relative error ($\epsilon_{criteria} = 0.04$) are used for ECS.

Figure 22 presents the prediction results for the oil production rates at P4 when the prior data include well responses close to observed data (case (i)). The true well response lies within the uncertainty intervals of the prediction results generated by DSI, LDFA, and LPFA. Conversely, Fig. 23 shows the prediction results for the oil production rates at P5 when the prior data do not include well responses close to the observed data (case (ii)). In this scenario, only LPFA provides an accurate uncertainty interval containing the true well response, as observed in the Brugge model.

The Brugge and Olympus models used in this study do not represent all geological settings. However, the Brugge model assumed three distinct depositional environments and was generated using different geostatistical techniques. Furthermore, the Olympus model was generated using standard geostatistical techniques for five facies. Having been successfully applied to complicated geological models such as the Brugge and Olympus models, it is expected that LPFA can be applied in different types of geological settings.

Conclusions

In this study, the learning-based pattern-driven forecast approach (LPFA) and ensemble conditioning step (ECS) are proposed to address the extrapolation, conditioning and prediction variance issues in data-space inversion (DSI) and the learning-based data-driven forecast approach (LDFA).

1. Accurate predictions can be achieved using LPFA, even when observed data are outside the prior data range. This capability arises from the transformation of prior data into pattern data using the new scaling method.
2. The application of ECS ensures that predictions sufficiently honor observed data while simultaneously reducing prediction variance. This outcome is achieved by selecting only learning models that provide predictions honoring the observed data.
3. Both the new scaling method and ECS are adaptable to a variety of data-driven forecast methods, helping to mitigate challenges associated with extrapolation, conditioning, and prediction variance.
4. The successful application of LPFA and ECS to complex geological models such as the Brugge and Olympus models suggests their potential effectiveness in diverse geological settings.
5. Significant advantages are offered by LPFA and ECS in terms of prediction accuracy and uncertainty reduction. However, as these methods are data-driven, retraining

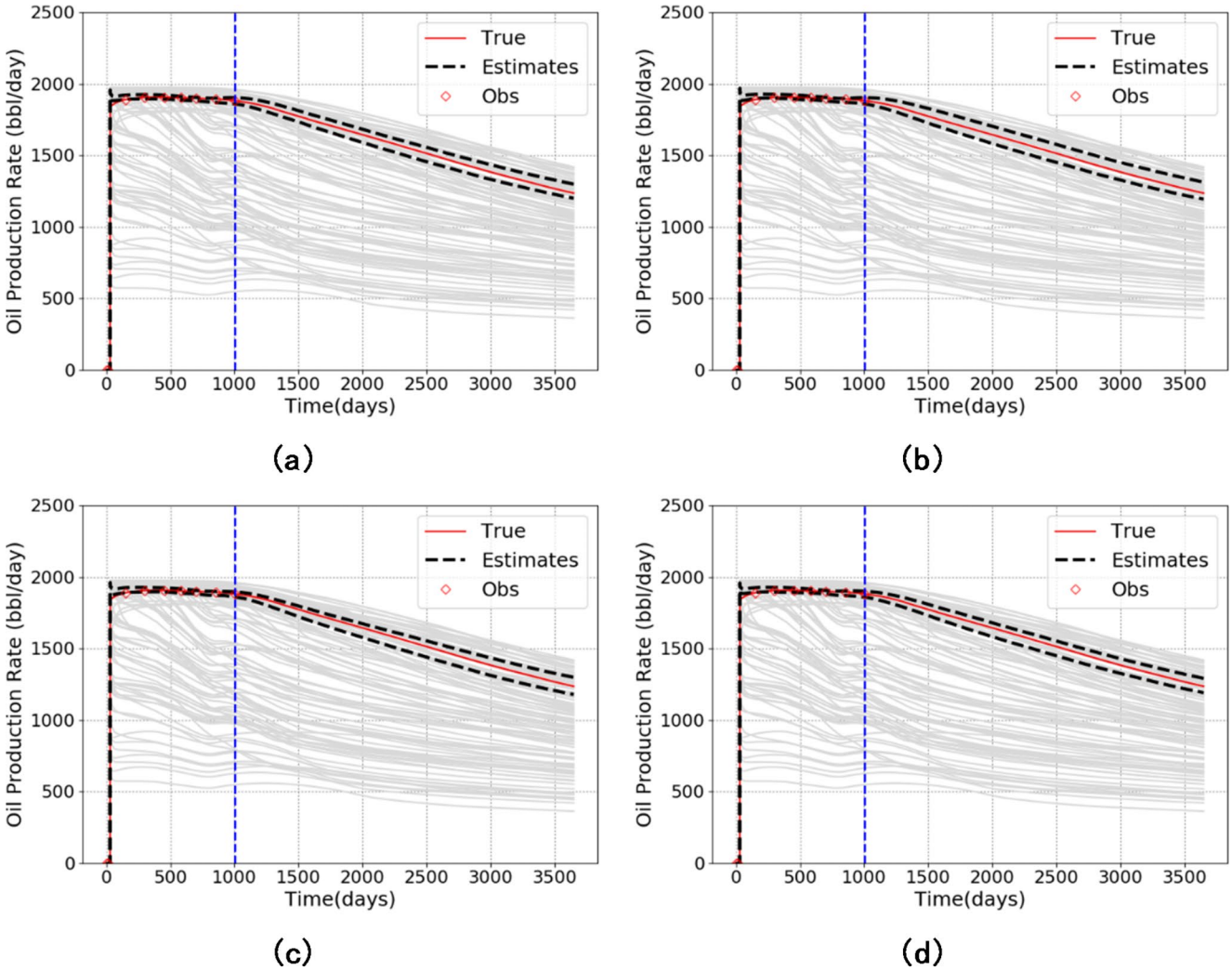


Fig. 18 Four predictions of P10 oil production rate using LPFA with ECS. Gray lines represent the well responses of prior models, and blue dashed line marks the end of the observation times

the learning model becomes necessary when operating conditions or geological information change. Moreover, addressing high uncertainty in reservoir parameters may require a large suite of models, resulting in higher computational costs. Balancing prediction accuracy with computational efficiency is therefore crucial for practical implementation. A potential solution to mitigate these computational challenges is to selectively and appropriately choose prior models that effectively represent response patterns, thereby reducing the number of simulation runs needed to generate prior data.

6. As data-driven methods, LPFA and ECS can be utilized to forecast various well responses, including water or gas injection rates and heat extraction rates. This adaptability makes them applicable to a wide range of subsurface applications, such as enhanced oil recovery, carbon capture and storage, and geothermal energy production. Their flexibility to integrate seamlessly with existing data-driven frameworks and adapt to diverse geological settings underscores their practical utility.

Appendix

To illustrate the new scaling method in detail, an example predicting the oil production rate of P9 in the Brugge model is provided. The Brugge model consists of 104 realizations, with the 104th realization selected as the true model for this example. Table 5 presents the oil production rates over time for both the training and test datasets. The oil production rates at the observation times ($\tau_1,\ldots,\tau_{N_\tau}$) for each realization serve as input to the learning model, while the rates at the forecast times ($t_1,\ldots,t_{N_t}$) serve as output.

As explained in the Sect. 3.1.1, each well response includes oil production rates at 498 dates ($N_t = 498$) spanning 0 to 3,648 days ($t_1, t_2,\ldots, t_{498} = 0, 1,\ldots, 3648$ days).

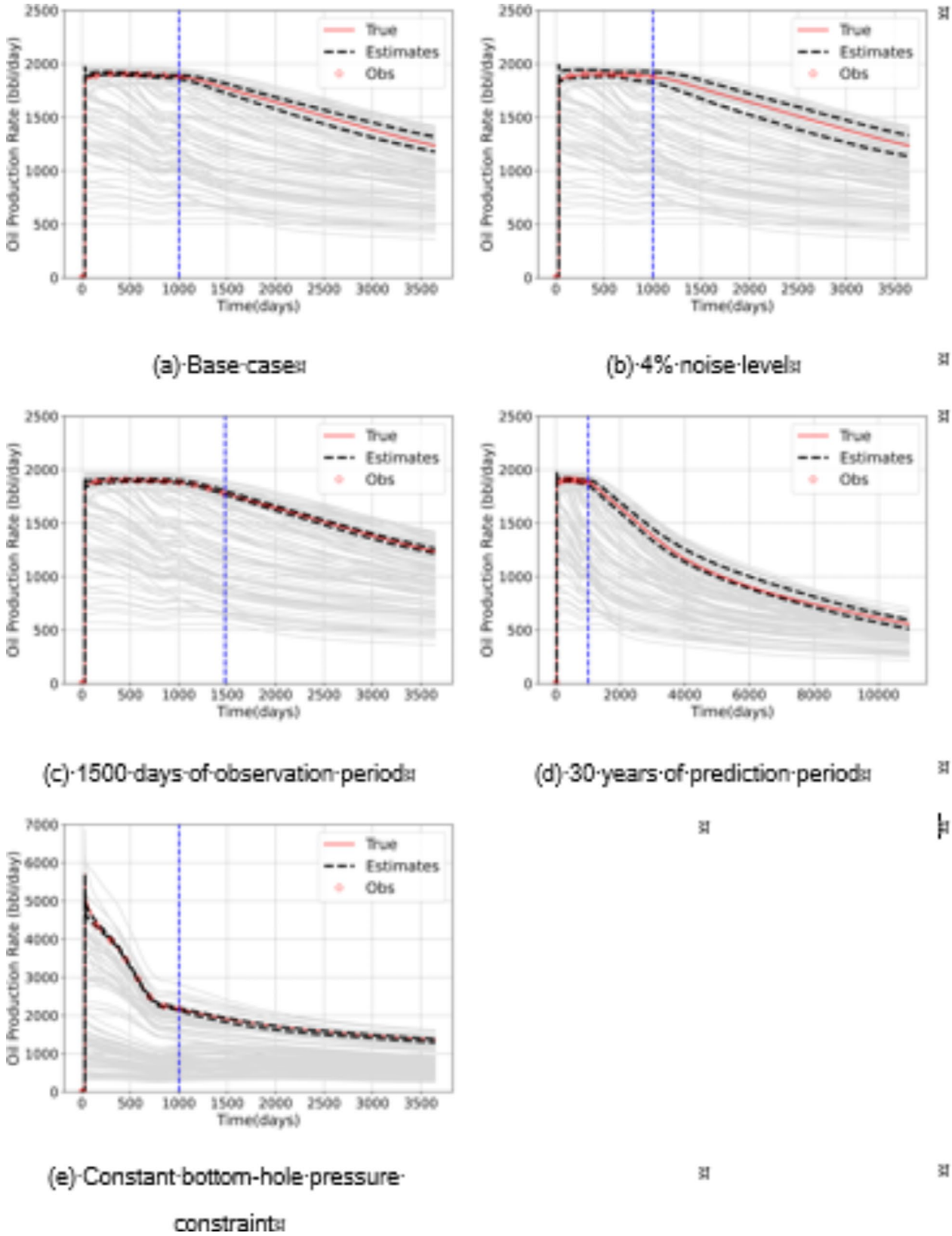


Fig. 19 Results of sensitivity analysis. Prediction results of P10 oil production rate (a) of base case (b) with a 4% noise level, (c) using observations up to 1500 days of observation period, (d) for 30 years of prediction period, and (e) under a constant bottom-hole pressure constraint. Gray lines represent the well responses of prior models, and blue dashed line marks the end of the observation times

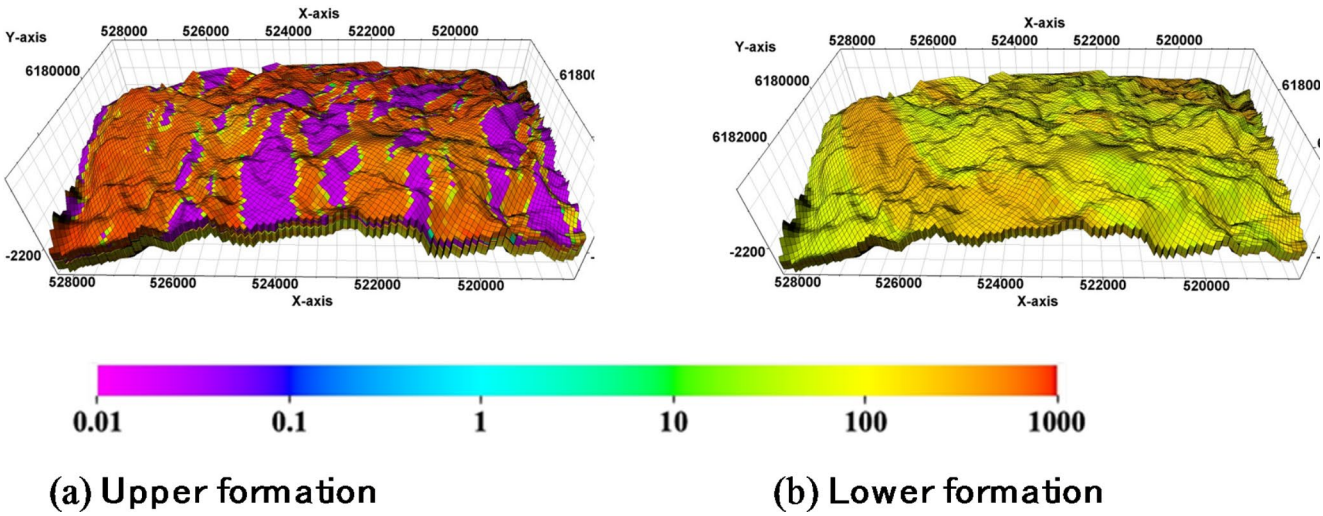


Fig. 20 Logarithm values of the x-direction permeability of a realization in the Olympus model set. The Olympus model has two rock formations: (a) the upper formation and (b) the lower formation, separated by an impermeable shale formation

Table 4 Reservoir properties for the different facies of the Olympus model

Facies type	Channel sand	Shale	Coarse sand	Medium sand	Fine sand
Permeability (md)	400 ~ 1,000	1	150 ~ 400	75 ~ 150	10 ~ 50
Porosity (%)	0.2 ~ 0.35	0.03	0.2 ~ 0.3	0.1 ~ 0.2	0.05 ~ 0.1
Net-to-gross (md)	0.8 ~ 1	0	0.7 ~ 0.9	0.75 ~ 0.95	0.9 ~ 1
Zone present	Upper formation	Upper formation, Barrier	Lower formation	Lower formation	Lower formation

Among the 498 dates, only 8 dates ($N_\tau = 8$, $\tau_1, \tau_2, \ldots, \tau_8 = 16, 151, 301, 451, 580, 703, 852, 1004$ days) are used as observation times.

The new scaling method scales data using the mean and standard deviation of each realization. When predicting a well response, only data from the observation period is available. Thus, the mean and standard deviation are calculated solely from the observed data. For the l-th realization, the input and output data are denoted as $d_{in,l}$ and $d_{out,l}$, respectively, while the mean and standard deviation of the input data are denoted as $\bar{d}_{\tau,l}$ and $\sigma_{\tau,l}$, respectively. As shown in the rightmost two columns of Table 5, the mean and standard deviation of the input data are computed for each realization.

Each realization is scaled using its own mean and standard deviation of the input data. For instance, the input and output data from the first realization are scaled using a mean of 103 and a standard deviation of 117. Similarly, the input and output data of the 104th realization are scaled using a mean of 18.9 and a standard deviation of 21.3. The results of the new scaling method are presented in Table 6. As illustrated in Fig. 4(c), the new scaling method enables the scaled test data lies within the distribution of the scaled training data, even when the test data falls outside the range of the training data.

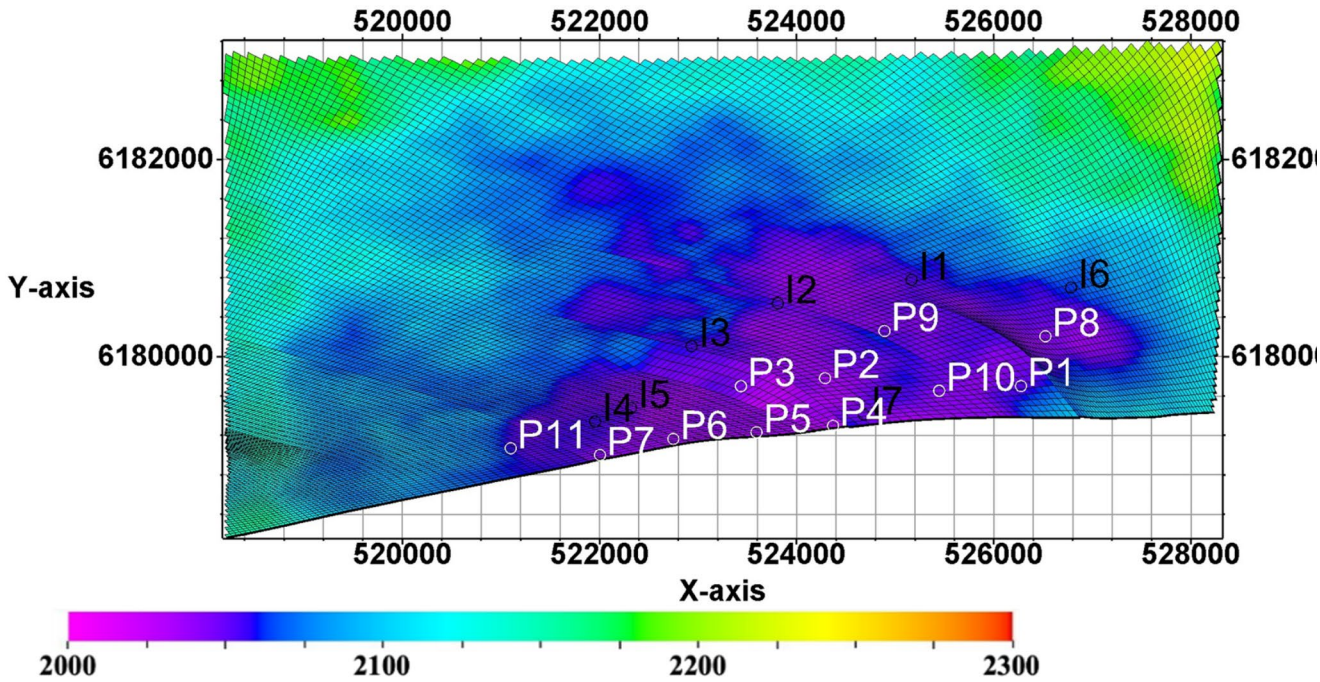


Fig. 21 Top depth (m) of the Olympus model and the location of the wells. Injectors and producers are labeled in black and white with I and P prefixes, respectively

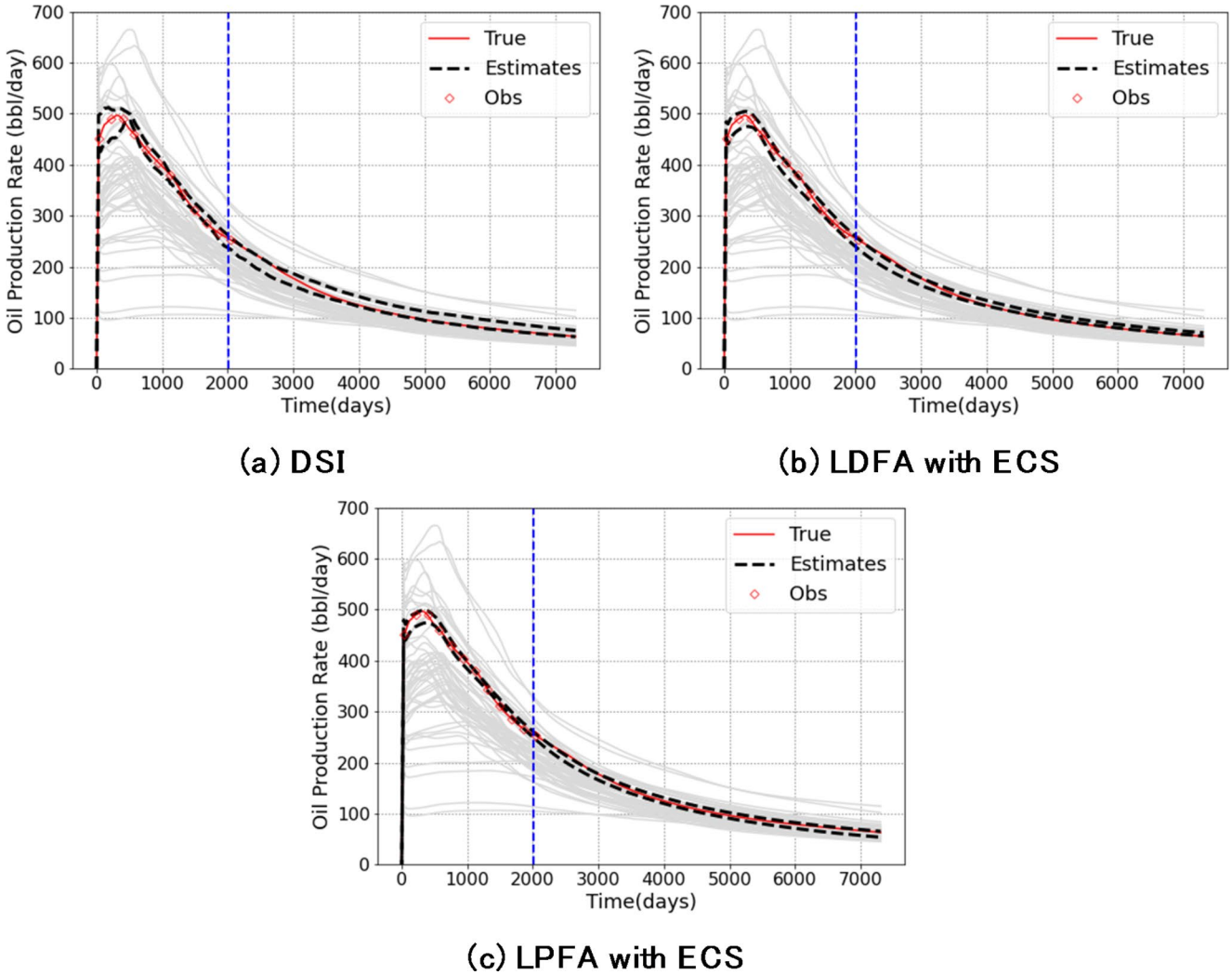


Fig. 22 Prediction results of P4 oil production rate using DSI, LDFA, and LPFA. Gray lines represent the well responses of prior models, and blue dashed line marks the end of the observation times

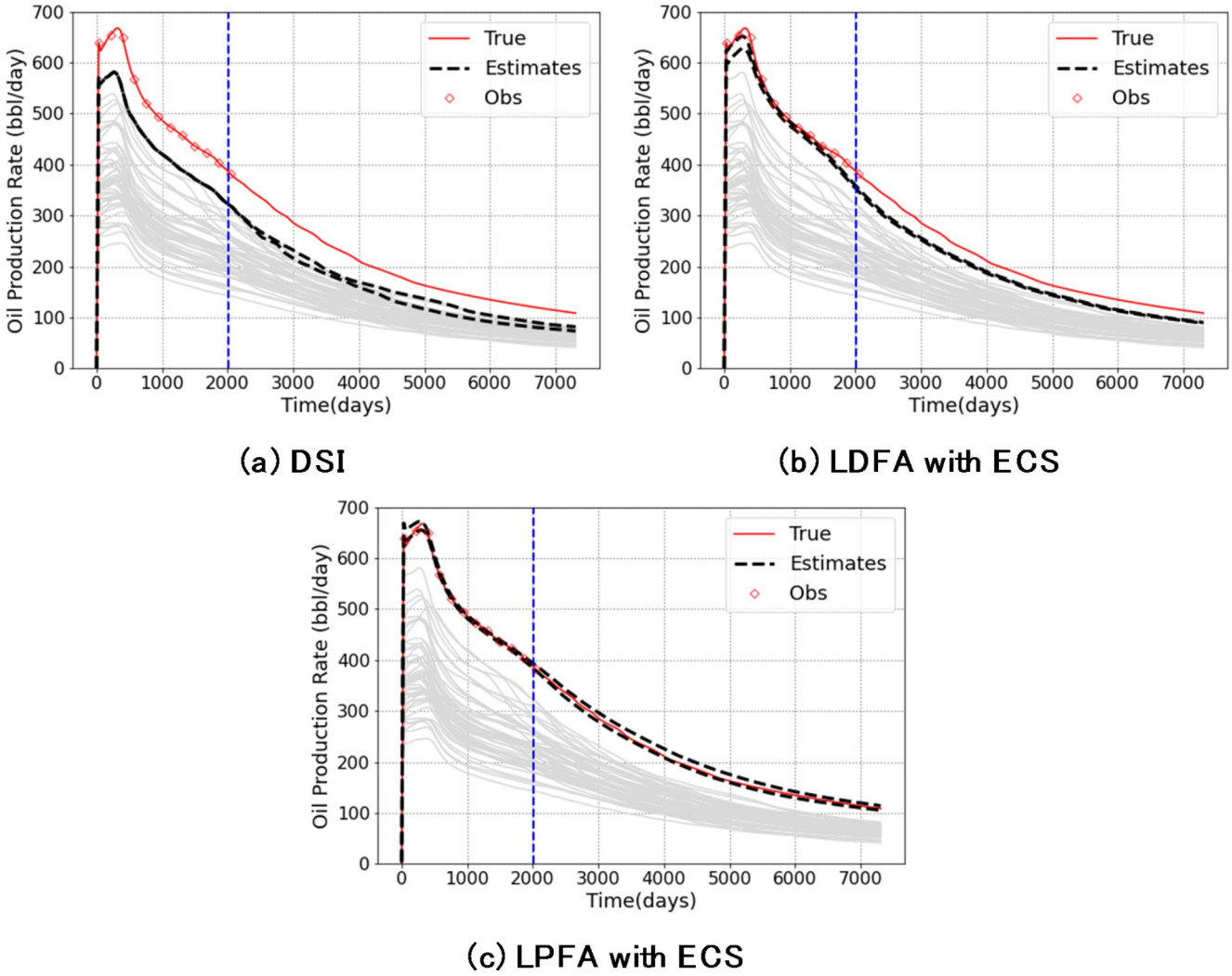


Fig. 23 Prediction results of P5 oil production rate using DSI, LDFA, and LPFA. Gray lines represent the well responses of prior models, and blue dashed line marks the end of the observation times

Table 5 P9 oil production rate over time for both training and test data sets. The rightmost two columns show the mean and standard deviation of input data for each realization

		Observation times (τ)	τ_1	τ_2	$\cdots$	τ_{N_τ}	-	$\cdots$	-	$\bar{d}_{\tau,l}$	$\sigma_{\tau,l}$
		Forecast times (t)	t_1	t_{10}	$\cdots$	t_{150}	t_{151}	$\cdots$	t_{N_t}		
Training data	**Realization 1**	Input ($d_{in,1}$)	0	0	$\cdots$	243	-	$\cdots$	-	103	117
		Output ($d_{out,1}$)	0	0	$\cdots$	243	245	$\cdots$	289	-	-
	Realization 2	Input ($d_{in,2}$)	0	0	$\cdots$	411	-	$\cdots$	-	160	180
		Output ($d_{out,2}$)	0	0	$\cdots$	411	415	$\cdots$	551	-	-
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
	Realization 103	Input ($d_{in,103}$)	0	0	$\cdots$	4.51	-	$\cdots$	-	2.61	3.06
		Output ($d_{out,103}$)	0	0	$\cdots$	4.51	4.48	$\cdots$	4.51	-	-
Test data	**Realization 104**	Input ($d_{in,104}$)	0	0	$\cdots$	39.2	-	$\cdots$	-	18.9	21.3
		Output ($d_{out,104}$)	0	0	$\cdots$	39.2	39.2	$\cdots$	50.1	-	-

Table 6 The results of the new scaling method. Each realization is scaled using its mean and standard deviation of the input data

		Observation times (τ)	τ_1	τ_2	$\cdots$	τ_{N_τ}	-	$\cdots$	-
		Forecast times (t)	t_1	t_{10}	$\cdots$	t_{150}	t_{151}	$\cdots$	t_{N_t}
Scaled train data	**Realization 1**	Input ($d_{in,\ scaled,1}$)	−0.879	−0.879	$\cdots$	1.19	-	$\cdots$	-
		Output ($d_{out,scaled,1}$)	−0.879	−0.879	$\cdots$	1.19	1.21	$\cdots$	1.59
	Realization 2	Input ($d_{in,scaled,2}$)	−0.887	−0.887	$\cdots$	1.40	-	$\cdots$	-
		Output ($d_{out,scaled,2}$)	−0.887	−0.887	$\cdots$	1.40	1.42	$\cdots$	2.17
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	Realization 103	Input ($d_{in,scaled,103}$)	−0.852	−0.852	$\cdots$	0.622	-	$\cdots$	-
		Output ($d_{out,scaled,103}$)	−0.852	−0.852	$\cdots$	0.622	0.611	$\cdots$	0.623
Scaled test data	**Realization 104**	Input ($d_{in,scaled,104}$)	−0.886	−0.886	$\cdots$	0.951	-	$\cdots$	-
		Output ($d_{out,scaled,104}$)	−0.886	−0.886	$\cdots$	0.951	0.952	$\cdots$	1.46

Acknowledgements We thank Schlumberger for providing ECLIPSE 2021.3.

Funding This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) grant funded by the Ministry of Trade, Industry and Energy (MOTIE) (No. 20214000000500, Training program of CCUS for the green growth) and Energy & Mineral Resources Development Association of Korea (EMRD) grant funded by MOTIE (No. 2021060002, Graduate educational program for Digital Oil Field).

Declarations

Conflict of interest The authors declare that there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless

indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Ahmed H, Lofstead J (2022) Managing Randomness to Enable Reproducible Machine Learning. Proceedings of the 5th International Workshop on Practical Reproducible Evaluation of Computer Systems. https://doi.org/10.1145/3526062.3536353

Bayes T (1763) LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. Philos Trans R Soc Lond 53:370–418. https://doi.org/10.1098/rstl.1763.0053

Bertolini AC, Schiozer DJ (2011) Influence of the objective function in the history matching process. J Pet Sci Eng 78:32–41. https://doi.org/10.1016/j.petrol.2011.04.012

Bigdeli A, Delshad M (2024) The Evolving Landscape of Oil and Gas chemicals: convergence of Artificial Intelligence and Chemical-enhanced oil recovery in the Energy Transition toward Sustainable Energy Systems and Net-Zero emissions. J Data Sci Intell Syst 2:65–78. https://doi.org/10.47852/bonviewJDSIS42022111

Cao J, Kitanidis PK (1999) Adaptive-grid simulation of groundwater flow in heterogeneous aquifers. Adv Water Resour 22:681–696. https://doi.org/10.1016/S0309-1708(98)00047-5

Chahar J, Verma J, Vyas D, Goyal M (2022) Data-driven approach for hydrocarbon production forecasting using machine learning techniques. J Pet Sci Eng 217:110757. https://doi.org/10.1016/j.petrol.2022.110757

Chavent G, Dupuy M, Lemmonier P (1975) History matching by Use of Optimal Theory. Soc Petrol Eng J 15:74–86. https://doi.org/10.2118/4627-PA

Chen Y, Oliver DS (2013) Levenberg–Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. Comput Geosci 17:689–703. https://doi.org/10.1007/s10596-013-9351-5

Chen WH, Gavalas GR, Seinfeld JH, Wasserman ML (1974) A New Algorithm for Automatic History Matching. Soc Petrol Eng J 14:593–608. https://doi.org/10.2118/4545-PA

Dietterich TG (2000) Ensemble methods in machine learning. In: Josef K and Fabio R (eds) Multiple Classifier systems. MCS 2000. Lecture notes in Computer Science. Springer, Berlin, Heidelberg, pp 1–15.

Dong X, Yu Z, Cao W et al (2020) A survey on ensemble learning. Front Comput Sci 14:241–258. https://doi.org/10.1007/s11704-019-8208-z

Ebrahimi F, Sahimi M (2004) Multiresolution Wavelet Scale up of unstable miscible displacements in Flow through Heterogeneous Porous Media. Transp Porous Media 57:75–102. https://doi.org/10.1023/B:TIPM.0000032742.05517.06

Emerick AA, Reynolds AC (2013) Ensemble smoother with multiple data assimilation. Comput Geosci 55:3–15. https://doi.org/10.1016/j.cageo.2012.03.011

Fonseca RM, Rossa E, Della, Emerick AA et al (2020) Introduction to the special issue: overview of OLYMPUS optimization Benchmark Challenge. Comput Geosci 24:1933–1941. https://doi.org/10.1007/s10596-020-10003-4

Gao G, Li G, Reynolds AC (2007) A stochastic optimization Algorithm for Automatic History Matching. SPE J 12:196–208. https://doi.org/10.2118/90065-PA

Gautier Y, Blunt MJ, Christie MA (1999) Nested Gridding and streamline-based Simulation for fast Reservoir Performance Prediction. Comput Geosci 3:295–320. https://doi.org/10.1023/A:1011535210857

Glorot X, Bengio Y (2010) Understanding the Difficulty of Training Deep Feedforward Neural Networks. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, PMLR. http://proceedings.mlr.press/v9/glorot10a.html. Accessed 19 Sep 2024

Goodfellow I, Bengio Y, Courville A (2016) Deep learning. MIT Press

Hassanpour RM, Manchuk JG, Leuangthong O, Deutsch CV (2010) Calculation of permeability tensors for unstructured Gridblocks. J Can Pet Technol 49:65–74. https://doi.org/10.2118/141305-PA

He C, Durlofsky LJ (2006) Structured flow-based gridding and upscaling for modeling subsurface flow. Adv Water Resour 29:1876–1892. https://doi.org/10.1016/j.advwatres.2005.12.012

Jeong H, Sun AY, Lee J, Min B (2018) A learning-based data-driven forecast approach for predicting future reservoir performance. Adv Water Resour 118:95–109. https://doi.org/10.1016/j.advwatres.2018.05.015

Jiang S, Durlofsky LJ (2021) Data-space inversion using a recurrent autoencoder for time-series parameterization. Comput Geosci 25:411–432. https://doi.org/10.1007/s10596-020-10014-1

Jiang S, Sun W, Durlofsky LJ (2020) A data-space inversion procedure for well control optimization and closed-loop reservoir management. Comput Geosci 24:361–379. https://doi.org/10.1007/s10596-019-09853-4

Jo S, Jeong H, Min B et al (2022) Efficient deep-learning-based history matching for fluvial channel reservoirs. J Pet Sci Eng 208:109247. https://doi.org/10.1016/j.petrol.2021.109247

Jolliffe IT (2002) Principal component analysis, 2nd edn. Springer, New York, NY, New York

Jung H, Jo H, Kim S et al (2020) Use of Channel Information Update and Discrete Cosine Transform in Ensemble Smoother for Channel Reservoir Characterization. J Energy Resour Technol 142:012901. https://doi.org/10.1115/1.4043856

Kang B, Jung H, Jeong H, Choe J (2020) Characterization of three-dimensional channel reservoirs using ensemble Kalman filter assisted by principal component analysis. Pet Sci 17:182–195. https://doi.org/10.1007/s12182-019-00362-8

Kim S, Min B, Lee K, Jeong H (2018) Integration of an iterative update of sparse geologic dictionaries with ES-MDA for history matching of Channelized reservoirs. Geofluids 2018:1–21. https://doi.org/10.1155/2018/1532868

Kim S, Lee K, Lim J et al (2020) Development of ensemble smoother–neural network and its application to history matching of channelized reservoirs. J Pet Sci Eng 191:107159. https://doi.org/10.1016/j.petrol.2020.107159

Lee K, Lim J, Yoon D, Jung H (2019) Prediction of shale-gas production at duvernay formation using deep-learning algorithm. SPE J 24:2423–2437. https://doi.org/10.2118/195698-PA

Li R, Reynolds AC, Oliver DS (2003) History matching of three-phase Flow Production Data. SPE J 8:328–340. https://doi.org/10.2118/87336-PA

Liang B, Liu J, You J et al (2023) Hydrocarbon production dynamics forecasting using machine learning: a state-of-the-art review. Fuel 337:127067. https://doi.org/10.1016/j.fuel.2022.127067

Lima MM, Emerick AA, Ortiz CEP (2020) Data-space inversion with ensemble smoother. Comput Geosci 24:1179–1200. https://doi.org/10.1007/s10596-020-09933-w

Liu D, Rao X, Zhao H et al (2021) An improved data space inversion method to predict reservoir state fields via observed production data. Pet Sci 18:1127–1142. https://doi.org/10.1016/j.petsci.2021.07.008

Liu P, Zhang K, Yao J (2023a) Reservoir automatic history matching: methods, challenges, and future directions. Adv Geo-Energy Res 7:136–140. https://doi.org/10.46690/ager.2023.02.07

Liu W, Xu Y, Rao X et al (2023b) Reservoir closed-loop optimization method based on connection elements and data space inversion with variable controls. Phys Fluids 35:113322. https://doi.org/10.1063/5.0172378

Luo X, Stordal AS, Lorentzen RJ, Nævdal G (2015) Iterative ensemble smoother as an approximate solution to a regularized minimum-average-cost problem: theory and applications. SPE J 20:962–982. https://doi.org/10.2118/176023-PA

Nwachukwu A, Jeong H, Pyrcz M, Lake LW (2018) Fast evaluation of well placements in heterogeneous reservoir models using machine learning. J Pet Sci Eng 163:463–475. https://doi.org/10.1016/j.petrol.2018.01.019

Oliver DS, Reynolds AC, Liu N (2008) Inverse theory for Petroleum Reservoir characterization and history matching. Cambridge University Press

Peters E, Arts RJ, Brouwer GK et al (2010) Results of the Brugge Benchmark study for flooding optimization and history matching. SPE Reservoir Eval Eng 13:391–405. https://doi.org/10.2118/119094-PA

Rwechungura R, Dadashpour M, Kleppe J (2011) Advanced History matching techniques reviewed. SPE Middle East Oil Gas Show Conf. https://doi.org/10.2118/142497-MS

Sagi O, Rokach L (2018) Ensemble learning: a survey. WIREs Data Min Knowl Discov 8:e1249. https://doi.org/10.1002/widm.1249

Scardapane S, Wang D (2017) Randomness in neural networks: an overview. WIREs Data Min Knowl Discov 7:e1200. https://doi.org/10.1002/widm.1200

Schlumberger (2021) ECLIPSE reference manual

Sun W, Durlofsky LJ (2017) A New Data-Space Inversion Procedure for efficient uncertainty quantification in Subsurface Flow problems. Math Geosci 49:679–715. https://doi.org/10.1007/s11004-016-9672-8

Sun W, Hui MH, Durlofsky LJ (2017) Production forecasting and uncertainty quantification for naturally fractured reservoirs using a new data-space inversion procedure. Comput Geosci 21:1443–1458. https://doi.org/10.1007/s10596-017-9633-4

Wang S, Qin C, Feng Q et al (2021) A framework for predicting the production performance of unconventional resources using deep learning. Appl Energy 295:117016. https://doi.org/10.1016/j.apenergy.2021.117016

Xiao C, Wang G, Zhang Y, Deng Y (2022) Machine-learning-based well production prediction under geological and hydraulic fracture parameters uncertainty for unconventional shale gas reservoirs. J Nat Gas Sci Eng 106:104762. https://doi.org/10.1016/j.jngse.2022.104762

Xiao C, Zhang S, Ma X et al (2023a) Robust production forecast and uncertainty quantification for waterflooding reservoir using hybrid recurrent auto-encoder and long short-term memory neural network. Geoenergy Sci Eng 230:212252. https://doi.org/10.1016/j.geoen.2023.212252

Xiao C, Zhang S, Ma X et al (2023b) Deep-learning-generalized data-space inversion and uncertainty quantification framework for accelerating geological CO2 plume migration monitoring. Geoenergy Sci Eng 224:211627. https://doi.org/10.1016/j.geoen.2023.211627

Zhong Z, Sun AY, Jeong H (2019) Predicting CO2 Plume Migration in Heterogeneous formations using conditional deep Convolutional Generative Adversarial Network. Water Resour Res 55:5830–5851. https://doi.org/10.1029/2018WR024592

Zhong Z, Sun AY, Wang Y, Ren B (2020) Predicting field production rates for waterflooding using a machine learning-based proxy model. J Pet Sci Eng 194:107574. https://doi.org/10.1016/j.petrol.2020.107574

Zhong Z, Sun AY, Ren B, Wang Y (2021) A deep-learning-based Approach for Reservoir Production Forecast under uncertainty. SPE J 26:1314–1340. https://doi.org/10.2118/205000-PA

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.