



# Characterization of three-dimensional channel reservoirs using ensemble Kalman filter assisted by principal component analysis

Byeongcheol Kang<sup>1</sup> · Hyungsik Jung<sup>1</sup> · Hoonyoung Jeong<sup>1</sup> · Jonggeun Choe<sup>1</sup>

Received: 21 December 2018  
© The Author(s) 2019

## Abstracts

Ensemble-based analyses are useful to compare equiprobable scenarios of the reservoir models. However, they require a large suite of reservoir models to cover high uncertainty in heterogeneous and complex reservoir models. For stable convergence in ensemble Kalman filter (EnKF), increasing ensemble size can be one of the solutions, but it causes high computational cost in large-scale reservoir systems. In this paper, we propose a preprocessing of good initial model selection to reduce the ensemble size, and then, EnKF is utilized to predict production performances stochastically. In the model selection scheme, representative models are chosen by using principal component analysis (PCA) and clustering analysis. The dimension of initial models is reduced using PCA, and the reduced models are grouped by clustering. Then, we choose and simulate representative models from the cluster groups to compare errors of production predictions with historical observation data. One representative model with the minimum error is considered as the best model, and we use the ensemble members near the best model in the cluster plane for applying EnKF. We demonstrate the proposed scheme for two 3D models that EnKF provides reliable assimilation results with much reduced computation time.

**Keywords** Channel reservoir characterization · Model selection scheme · Egg model · Principal component analysis (PCA) · Ensemble Kalman filter (EnKF) · History matching

## List of symbols

<b>C</b>	Covariance matrix
<b>H</b>	Measurement operator matrix
<b>K</b>	Kalman gain
<b>N</b>	Total number of ensemble members
<b>X</b>	Matrix of ensemble vectors
<b><math>\bar{X}</math></b>	Matrix of mean values of ensemble vectors
<b>Y</b>	Group of ensemble members
<b>x</b>	Vector contains reservoir properties of all grid blocks
<b>y</b>	Ensemble members

## Superscript

<b>a</b>	Assimilated
<b>p</b>	Prior

## Subscript

<b>y</b>	Ensemble members
<b>m</b>	Measurement error

## 1 Introduction

Integration of various reservoirs' information improves accuracy and reliability of reservoir models. However, the uncertainty associated with the reservoir models is likely to be significant, because very sparse static data such as core and well log data would be available to characterize a reservoir. These uncertainties can be reduced by calibrating reservoir parameters, so the responses of the reservoir model are matched to measurement data such as flow rates and bottom-hole pressures. Then, its future performances can be predicted reliably using the calibrated reservoir model honoring the measurement data.

Ensemble Kalman filter (EnKF) proposed by Evensen (1994) is one of the popular history matching (or model

Edited by Yan-Hua Sun

✉ Jonggeun Choe  
johnchoe@snu.ac.kr

Byeongcheol Kang  
qudcjf@snu.ac.kr

Hyungsik Jung  
hs6735@snu.ac.kr

Hoonyoung Jeong  
hoonyoung.jeong@snu.ac.kr

<sup>1</sup> Department of Energy Systems Engineering, Seoul National University, Seoul 08826, South Korea

calibration) methods because EnKF is not only easy to couple with any flow simulator, but also parallelizable and computationally efficient. EnKF has been applied in a variety of studies such as oceanography, meteorology, hydrology, and petroleum engineering (Kalman 1960; Houtekamer and Mitchell 2001; Nævdal et al. 2005; Liu and Oliver 2005; Jafarpour and McLaughlin 2009; Kang and Choe 2017). For history matching and uncertainty quantification in reservoir engineering studies, Nævdal et al. (2002) used EnKF for the first time to update the permeability distribution near the wellbore.

EnKF updates reservoir properties of interest using available measured data. Arroyo-Negrete et al. (2008) utilized streamlines and Jeong et al. (2010) added a gradual deformation method to EnKF so that they calculated Kalman gain with better reliability. Lee et al. (2013) used clustering analysis and they made several clusters by considering permeability distributions. In each cluster, the reservoir models are updated using different Kalman gains. There are also some schemes to update the model parameters selectively in EnKF. It is called as the localization scheme, and these schemes are divided into distance-based (Chen and Oliver 2010; Jung et al. 2017b), attribute-based (e.g., streamline and temperature distribution) (Arroyo-Negrete et al. 2008; Huang and Zeng 2016), and correlation-based process (Luo et al. 2018, 2019).

van Leeuwen and Evensen (1996) proposed ensemble smoother (ES), which utilizes a single global update at the end of observation times. EnKF updates an ensemble recursively, which means that reservoir properties are updated every time when we get new measurement data. Compared to the multiple updates of EnKF, ES conducts only one update by using all historical data available. For this reason, ES is faster than EnKF in updating the ensemble, but ES tends to give unstable results in highly complex problems. Thus, many researchers have attempted to improve Kalman gain estimations for stable convergence in ES (Kang et al. 2016; Lee et al. 2016, 2017).

It is mathematically proven that initial models affect quality of final models (Evensen 2004; Jafarpour and McLaughlin 2009; Kang and Choe 2017). However, as mentioned above, even though prior data such as core samples, well log data, and geological survey data are integrated into the initial (or prior) models, many equiprobable reservoir models are possible because the prior data are very sparse compared to the entire reservoir size. Therefore, the matching quality of historical and predicted responses can be improved by using initial models selectively in EnKF or ES.

Kim et al. (2017) proposed an initial ensemble design scheme that resamples initial ensemble members that exhibit similar fluid migrations. Jung et al. (2017b) suggested a different initial ensemble design scheme that selects a subset of initial ensemble models based on the relative errors

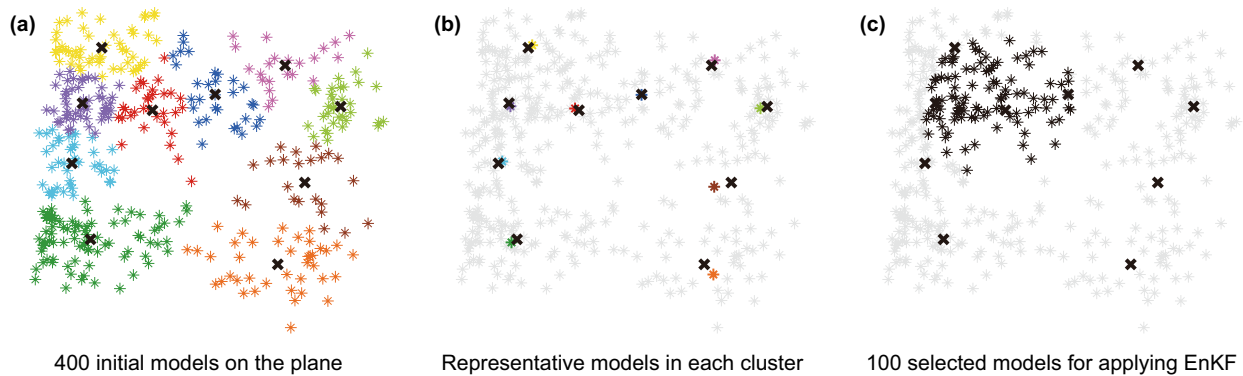
between historical observation data and predicted data. These schemes by Kim et al. (2017) and Jung et al. (2017b) improve the history matching quality for channel reservoirs, but they require pre-simulation runs for all initial ensemble members. Thus, the initial ensemble design schemes are computationally demanding for large-scale reservoir models.

We propose a model selection scheme using principal component analysis (PCA) and clustering analysis. PCA is a useful mathematical tool to manage high-dimensional data by extracting primary parameters of the data (Lim et al. 2015; Siena et al. 2016; Jung et al. 2018). Some researchers utilized PCA with other schemes to apply to multipoint geostatistics (Vo and Durlofsky 2014, 2015, 2016; Chen et al. 2016). We use PCA to extract the primary characteristics of channel distribution and select good models efficiently.

There are several ways to use PCA for data analysis. One is to apply PCA to reparametrize high-dimensional data into low-dimensional representation. Vo and Durlofsky (2014) generated new realizations based on principal parameters from PCA, and Vo and Durlofsky (2015, 2016) used the regenerated models for data assimilation. The reparametrization process reduces dimension while it maintains principal information of original data. Likewise, singular value decomposition and cross-validation scheme can be also used for reducing dimension of data (Sætrom et al. 2012; Patel et al. 2015; Kang et al. 2016).

Another approach is to manage the data in reduced-order dimension, considering principal characteristics by PCA. Our strategy is to choose reliable initial models, which have similar reservoir parameters before applying EnKF, by simulating only a few representative models. For the model selection, we group the initial models in a dimension-reduced space using PCA. The models in the same group will have similar reservoir characteristics of interest. Therefore, we can check the group's character by simulating just one center model. Only the representative models of the groups are simulated, and then, the mismatch is compared between observed data and simulated data. The best model that has the minimum mismatch is chosen. In this study, only 100 models that are close to the chosen model are selected as reliable models and updated further using EnKF.

Kang et al. (2017) also used PCA to select good initial models and they applied the selected models to EnKF. Their applications are limited to mostly 2D models. However, a 3D reservoir is much more complex than 2D because we should consider both vertical and horizontal reservoir parameters with increased number of unknown parameters. Although there have been many schemes proposed and demonstrated for 2D cases, only a few of them are applied for 3D reservoirs due to these difficulties. An application to 3D is not a simple extension of the dimension but a big task to manage these difficulties for fast and reliable results. In this paper, we demonstrate the efficiency of the PCA-assisted model



**Fig. 1** Dataset of 400 reservoir models projected on the 2D plane using PCA in this study. The clusters are denoted as different colors.  $\times$  is the centroid of each cluster

selection scheme in 3D channel reservoir models. It will save computational time without simulating all initial ensemble models, if the proposed scheme works for complex 3D channel reservoirs.

## 2 Methodology

### 2.1 The model selection process using PCA and clustering analysis

Channel reservoirs are highly heterogeneous because rock facies are clearly different. In common, sand provides flow channels while shale prevents flows. Because oil and gas flow along the sand channels, the connectivity of channels is a critical issue in forecasting future performances of the reservoir.

Models with similar channel trends will have similar production histories. Therefore, we use PCA to characterize reservoir parameters efficiently by grouping reservoir models based on channel distribution. Each grid has permeability values and we can construct a vector  $\mathbf{x}$ , which contains the properties of all grid blocks. We calculate eigenvalues and eigenvectors by decomposing the covariance matrix of all the models. The covariance matrix  $\mathbf{C}$  is computed using Eq. (1):

$$\mathbf{C} = \frac{1}{N-1} (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})^T \quad (1)$$

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \quad (2)$$

where  $N$  is the total number of the models and each column of  $\bar{\mathbf{X}}$  is the mean value of  $\mathbf{x}$ .

The magnitude of an eigenvalue implies the degree of the principal direction of the corresponding eigenvector. We use them to extract primary characteristics of the data. In this paper, we choose a subset of eigenvectors corresponding

to large eigenvalues and project the models onto the space formed using the eigenvectors. Because all eigenvectors are orthogonal, we can make a space using a few eigenvectors and project the models into the space of low dimensions.

Figure 1 is the 2D space obtained using two principal components of the 400 initial models in this study. The space is constructed by using the two eigenvectors, which are corresponding to the largest eigenvalues. The reservoir models are denoted as a dot in the plane, and the distances between the models show how dissimilar they are. This concept of ‘distance’ for comparing the dissimilarity has been employed in many studies to group models (Suzuki et al. 2008; Scheidt and Caers 2009; Kim et al. 2017; Jung et al. 2017b; Koneshloo et al. 2017). We measure the dissimilarity of the permeability distribution of the models to make groups according to their characteristics.

Some researchers used multi-dimensional scaling (MDS) to reduce the data’s dimensions. Each distance is required to construct a distance matrix for using MDS so that there are many proposed methods to define and measure difference between the data (Scheidt and Caers 2009; Chiotoroiu et al. 2017; Lee et al. 2017).

It is computationally inefficient if we simulate all ensemble models. Instead, we choose representative models of the groups obtained by PCA and clustering analysis. The initial models are divided by using  $K$ -means clustering (Fig. 1a), which is one of the efficient methods to make several groups. We set the models into ten clusters so that we compare the relative error of production data of the ten representative models (Fig. 1b). Near the model with the least error, we can select additional models in the plane for applying EnKF, since EnKF utilizes many equiprobable models. Finally, 100 models are selected in this case (Fig. 1c) and they have similar channel trends because we choose the models considering both static and dynamic reservoir data.

Table 1 provides overall workflow for selecting reservoir models. After generating initial reservoir models, we project

**Table 1** Workflow of the proposed model selection scheme

Work to do
1. Generate initial channel reservoir models
2. Calculate eigenvalues and eigenvectors
3. Construct a 2D space by using the two eigenvectors
4. Project the initial models on the plane
5. Divide the models into ten groups by using <i>K</i> -means clustering
6. Choose ten representative models to calculate the error of production data
7. Select 100 reservoir models near the model with the minimum error

them on the 2D space using PCA. The number of principal components to reduce the dimension depends on the complexity of data and user's objective. There is a mathematical formula to determine the reduced dimensions in PCA as Eq. (3):

$$\frac{\sum_{i=1}^p \lambda_i}{\sum_{i=1}^P \lambda_i} > \alpha \quad (3)$$

where  $\lambda$  is the eigenvalue,  $\alpha$  is the threshold value,  $P$  is the total number of eigenvalues, and  $p$  is the number of selected eigenvalues to reduce the dimension.  $\alpha$  is typically set to 0.80–0.95, but it depends on the data. In our case, it does not improve classification to maintain the principal components over 80%. Therefore, we used the 2 largest eigenvalues to denote the initial models on the reduced space.

The initial models are located according to their permeability distributions. Next, we find the best centroid model by comparing production profiles with measurement data. Finally, 100 reservoir models are selected near the best model on the PCA-based plane. The number of initial models, clusters, and selected models can be changed depending on each study. These settings should be determined on consideration of complexity of reservoir problem and computer hardware.

## 2.2 Ensemble Kalman filter

Ensemble Kalman filter (EnKF) has two main processes: prediction and assimilation steps. In the prediction step, forward reservoir simulation is conducted for its prediction until the next time of available observation data. Then, we update the models by using Kalman gain matrix in the assimilation step. The models can be repetitively updated when it comes to new observation data, so EnKF is useful for a real-time update procedure.

In EnKF, all reservoir models, called ensemble, are a group of state vectors  $y$  and they are constructed by reservoir variables and predicted data. We update permeability values in this paper, but other properties such as porosity, facies ratio, and aquifer strength can be put into the state vectors

and updated (Zhao et al. 2008; Jung et al. 2017a; Kim et al. 2017). In the assimilation step, ensemble  $Y$  is updated by Eq. (4), Kalman gain  $K$  and the difference between observed and predicted data. Kalman gain in Eq. (6) is derived to minimize the estimated error covariance  $C_Y$

$$Y^a = Y^p + K(d_i - HY^p) \quad (4)$$

$$Y = [y_1 \cdots y_{N_e}] \quad (5)$$

$$K = C_Y^p H^T (H C_Y^p H^T + C_m)^{-1} \quad (6)$$

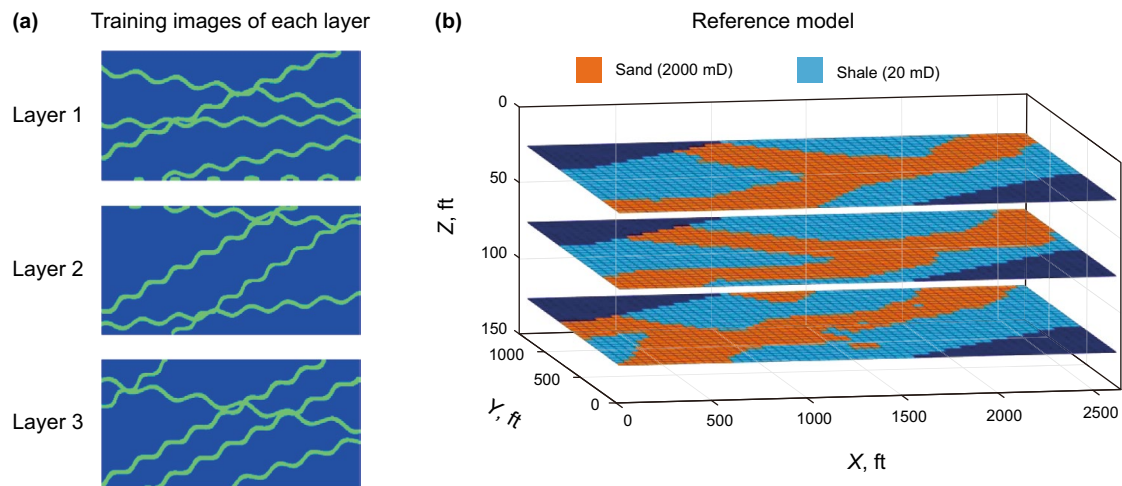
where  $C_m$  is the measurement error covariance.  $H$  is the measurement operator matrix in which entries are 0 or 1 only. Superscripts  $a$  and  $p$  denote 'assimilated' and 'prior,' respectively.

## 3 Results

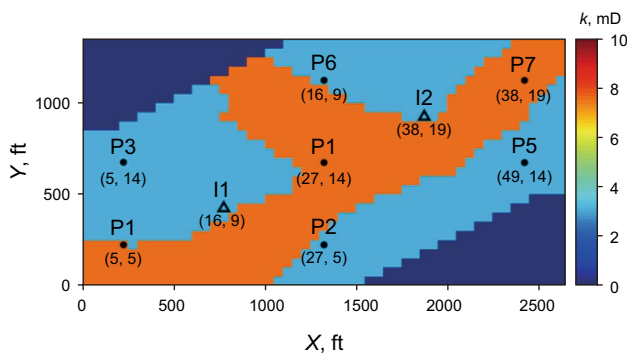
### 3.1 3D channel reservoir

We make a 3D channel reservoir using single normal equation simulation (SNESIM) module in SGeMS (Remy et al. 2011). The reservoir has a  $53 \times 27 \times 3$  grid system and each grid is a cubic with 50 ft long. We use three training images (TI) to generate diverse channel reservoir models. The TIs have multiple sinusoid curves of primarily upward to the right side (Fig. 2a). The reference model in this case is in Fig. 2b. We assume that the reservoir has two uniform permeability values of sand (2000 mD) and shale (20 mD). There are inactive cells on upper left and lower right side of the field, so 3513 grids are used for flow simulations.

The top view of the reference model is shown in Fig. 3. There are seven production wells and two injection wells on the field, and they are completed in all vertical layers. The total operation period is 2000 days, and it is assumed that we have produced oil and gas for 1000 days. The historical oil production rates at each well are given every 100 days, and they are used to update the reservoir models in EnKF. The



**Fig. 2** Information on the 3D channel reservoir in this study



**Fig. 3** Reference model from the top view (P: production well; I: injection well;  $k$ : log-permeability)

specification of the computer for the study is Intel i7-4770K with 3.50 GHz and 16.0 GB RAM.

The prediction quality of EnKF results is compared in three different cases. Case (a) uses all the 400 initial models in EnKF. Different model sampling methods are tested in the other cases: Case (b) uses the uniform sampling method and case (c) does the proposed selection scheme. In the cases (b) and (c), only 100 reservoir models are selected among the initial 400 models.

Since case (a) uses 4 times more ensemble than the other two cases, we intuitively expect stable results after EnKF updates in case (a). The mean of the permeability field after applying EnKF is shown in Fig. 4. All figures in Fig. 4 have same scales in log-transformed values. The color bar is on the right. The first column is the reference model and the second is one of the 400 initial models. For the initial models, the layer 1 is similar to that of the reference model. However, there are discontinuities of the channels in the layers 2 and 3, which need proper assimilations for reliable production estimation. Case (a) in the third column shows

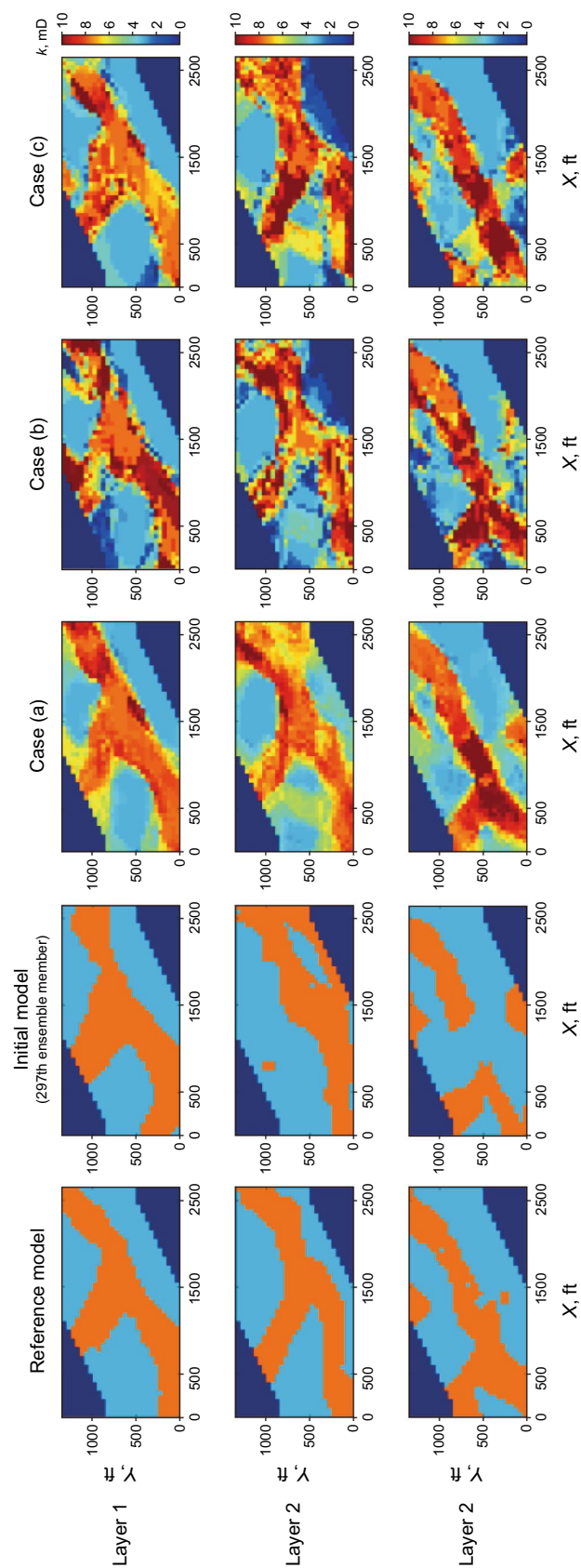
good updates in layers 1 and 3. Particularly, the channel in layer 3 is connected after the EnKF updates. However, in layer 2, permeability is overestimated along the boundary between sand and shale, and the channel to the left top is still not clear.

EnKF in case (b) looks to reproduce the channel connectivity but the permeability values are much higher than those of the reference. It is called as overshooting resulting from small ensemble size or unstable convergence of the Kalman gain. The overshooting problem leads to unreliable prediction of future productions, so it needs an additional remedy to manage the problem. For case (c) (the fifth column), there are high-permeability zones but the degree is not severe as case (b). In addition, the connectivity of the channels in case (c) is improved compared to case (b) to follow the reference's trends properly: connecting the channel in the northwest direction in the layer 2 and the diagonal channel in the layer 3 from disconnected channels in the initial model.

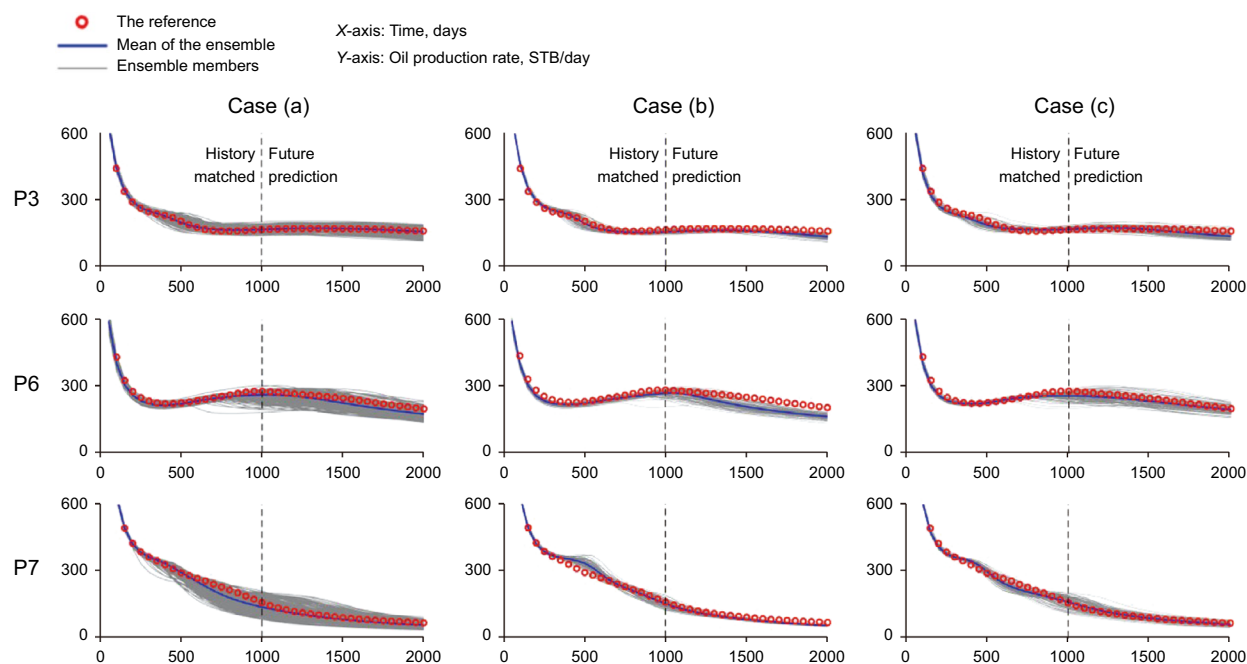
For characterizing channel reservoirs, it is important to figure out connectivity and direction of channels to predict production performances reliably. It is because fluid flows along sand facies of high permeability in channel reservoirs. Therefore, overall channel trend should be found out through the EnKF updates. For case (c) in Fig. 4, the channel connections are well calibrated so that we expect reliable estimations on production performances as shown later in Figs. 5, 6, and 7.

Table 2 shows L2 norm of the updated permeability values for quantitative comparison of the three cases. It is calculated by using log-transformed values. L2 norm indicates overall accuracy compared to the true model. Case (a) has the minimum value among the three cases. L2 norm in case (c) is larger than case (a), but smaller than the case using same ensemble size. Compared to case (b), the proposed method reduces the error by 14.13%.





**Fig. 4** Mean permeability fields after EnKF updates of the three cases ( $k$ : log-permeability)



**Fig. 5** Oil production rates in the production wells P3, P6, and P7 after EnKF updates of the three cases. The red dotted line is the reference and the thick blue line is the mean of the ensemble. Gray lines are the results of all reservoir models used

**Table 2** L2 norm of updated permeability values among the three cases

Case	Case (a)	Case (b)	Case (c)
L2 norm	1.859	2.725	2.340

The average of a large number of ensemble models will help to calibrate reservoir properties in EnKF. However, simulation cost for the 400 realizations is a burden and future predictions are not reliable (Fig. 7). Therefore, it is not practically recommended to use hundreds of models, especially in 3D models. Considering that case (c) uses a quarter number of ensemble models in case (a), it is a reasonable solution for both speed and reliability of the results.

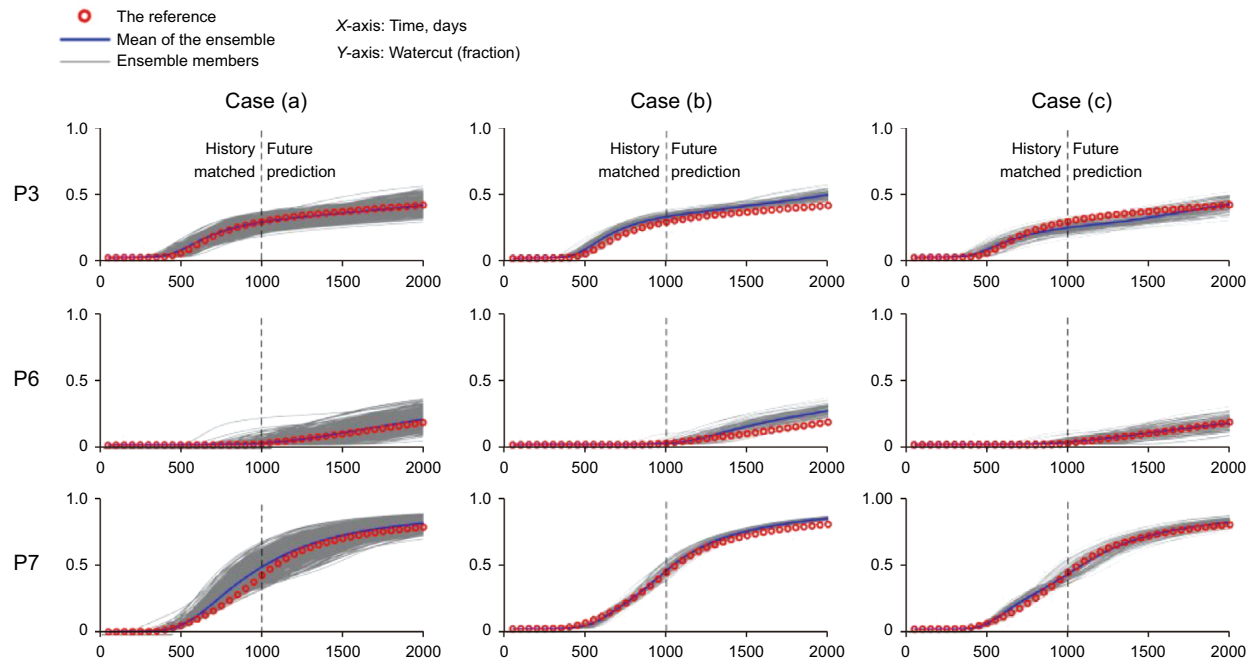
Note that it starts with the same initial models, but the EnKF results are different due to ensemble members used. It could be a safe guideline to use a large ensemble size for stable convergence of EnKF updates, but it is not necessarily the best considering simulation efficiency and reliability. Total simulation time for each case is summarized in Table 3. Case (a) spends almost 5.5 h for the whole process, whereas the other cases do 1.5 h. Model selection process takes only a few minutes so that ensemble size is a dominant factor on computing time. By using the proposed sampling scheme, 72.17% of the total simulation time is saved compared to case (a), and the connectivity of the channels is reliably reproduced.

**Table 3** Simulation time among the three cases

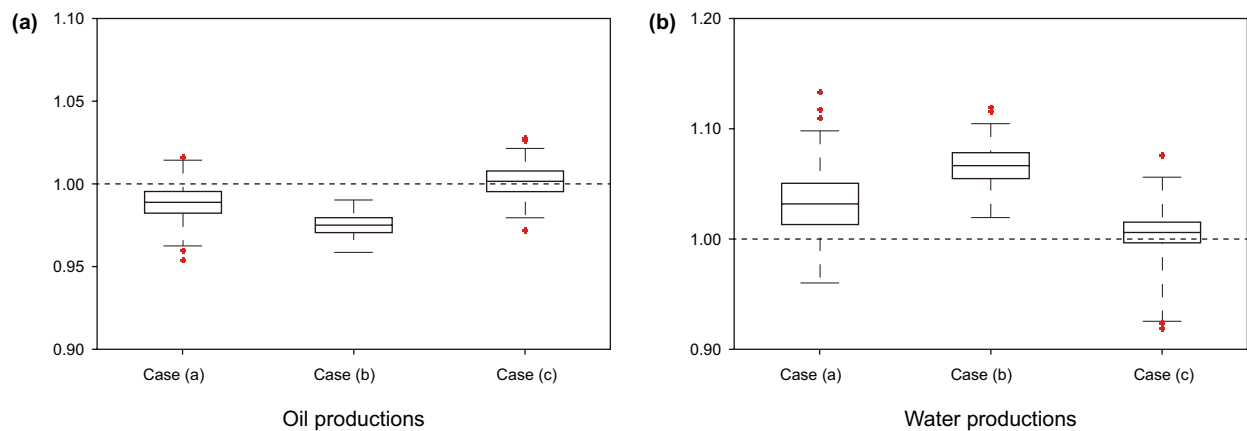
Case	Time, min	Time reduction, %
Case (a)	327	0
Case (b)	89	72.78
Case (c)	91	72.17

History matching results are shown in Figs. 5 and 6 on oil production rates and watercuts (water rate to total liquid rate) at production wells P3, P6, and P7 after the update using EnKF. The vertical dash line in the figures represents the last time of the observations. This means that we update the reservoir models until the 1000th day and we predict the next 1000 days. The mean values in case (b) are deviated from the reference, and the predictions by the ensemble members do not cover the true line. On the other hand, the proposed case shows much more reliable prediction than case (b) in both oil and water productions. The predictions in case (a) look stable, predicting the true production values, but the forecast range is much larger than that in the case (c). This is also confirmed from the boxplots in Fig. 7.

For quantitative comparisons of the three cases, we analyze the uncertainty range and the production forecasts of oil and water. At first, we calculate the difference between maximum and minimum value of oil rates and watercuts to measure prediction range of the ensemble in 50 days



**Fig. 6** Watercuts in the production wells P3, P6, and P7 after EnKF updates of the three cases



**Fig. 7** Boxplots on the cumulative productions after 2000 days

**Table 4** Reduction ratio (%) of prediction ranges compared to case (a)

Wells	Oil production rate			Watercuts		
	Case (a)	Case (b)	Case (c)	Case (a)	Case (b)	Case (c)
P3	0.000	42.430	17.976	0.000	43.596	13.546
P6	0.000	33.784	28.771	0.000	52.399	44.107
P7	0.000	57.118	38.247	0.000	68.169	49.210
Average	0.000	44.444	28.331	0.000	54.721	35.621



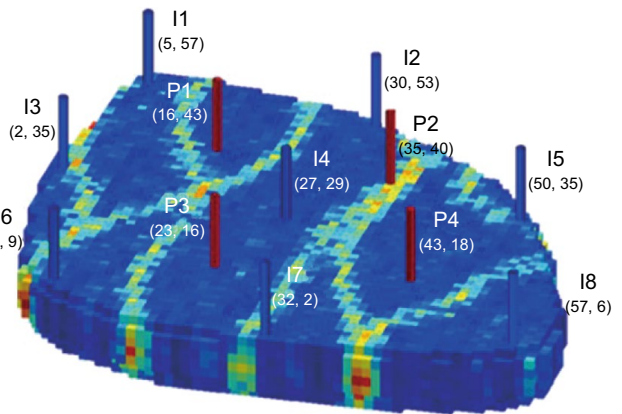
**Table 5** RMSE of oil production rates and watercuts

Wells	Oil production rates			Watercuts		
	Case (a)	Case (b)	Case (c)	Case (a)	Case (b)	Case (c)
P3	3.203	11.954	12.425	0.010	0.044	0.033
P6	18.196	29.649	14.218	0.001	0.040	0.003
P7	15.238	13.920	10.997	0.038	0.025	0.018
Average	12.212	18.508	12.547	0.020	0.037	0.018

interval. Table 4 shows the reduction ratio of uncertainty ranges of the three wells compared to the case (a). In both cases (b) and (c), the uncertainty ranges are reduced for oil production rates and watercuts. Although the average reduction ratio in case (b) is higher than that of case (c), it does not imply case (b) is better than case (c). Case (b) provides too narrow prediction ranges for uncertainty quantification. In many cases of variance underestimation, also known as filter divergence, the predicted results could not cover the true as seen in Figs. 5 and 6. In addition, it is impossible to estimate future productions stochastically as a negative effect.

Since the reduction ratio alone cannot give a clear clue for reliable results, we conduct another comparison by root-mean-square error (RMSE) between the true value and the mean of the ensemble. Table 5 shows the result of the RMSE analysis on oil and water productions in 50 days interval for 2000 days. Case (b) has the highest RMSE in average and most of the wells for oil and water productions. Although the uncertainty ranges are small and look stable, the accuracy of the production forecasts in case (b) is worse than the other cases. For well P7, RMSE is large in case (a) due to wide prediction ranges. In contrast, case (c) has smaller RMSE as well as it reduces prediction ranges compared to case (a). Therefore, the proposed method helps to improve the history matching results, considering both stability and accuracy.

The boxplot about the total productions is compared (Fig. 7) at the end of the operation period for quantitative comparison in another way. The values are normalized by the reference, so we draw a horizontal line crossing 1 to denote the true value. The boxes in case (b) do not cover the true value, which means wrong estimations on the oil and water productions. Also note that the medium of case (a) does not match with the true values for both oil and water productions. Case (c) covers the true production amounts in the box ranges, which is more accurate than the others. Therefore, the selection of good initial reservoir models helps to improve the prediction quality as well as the computational cost.

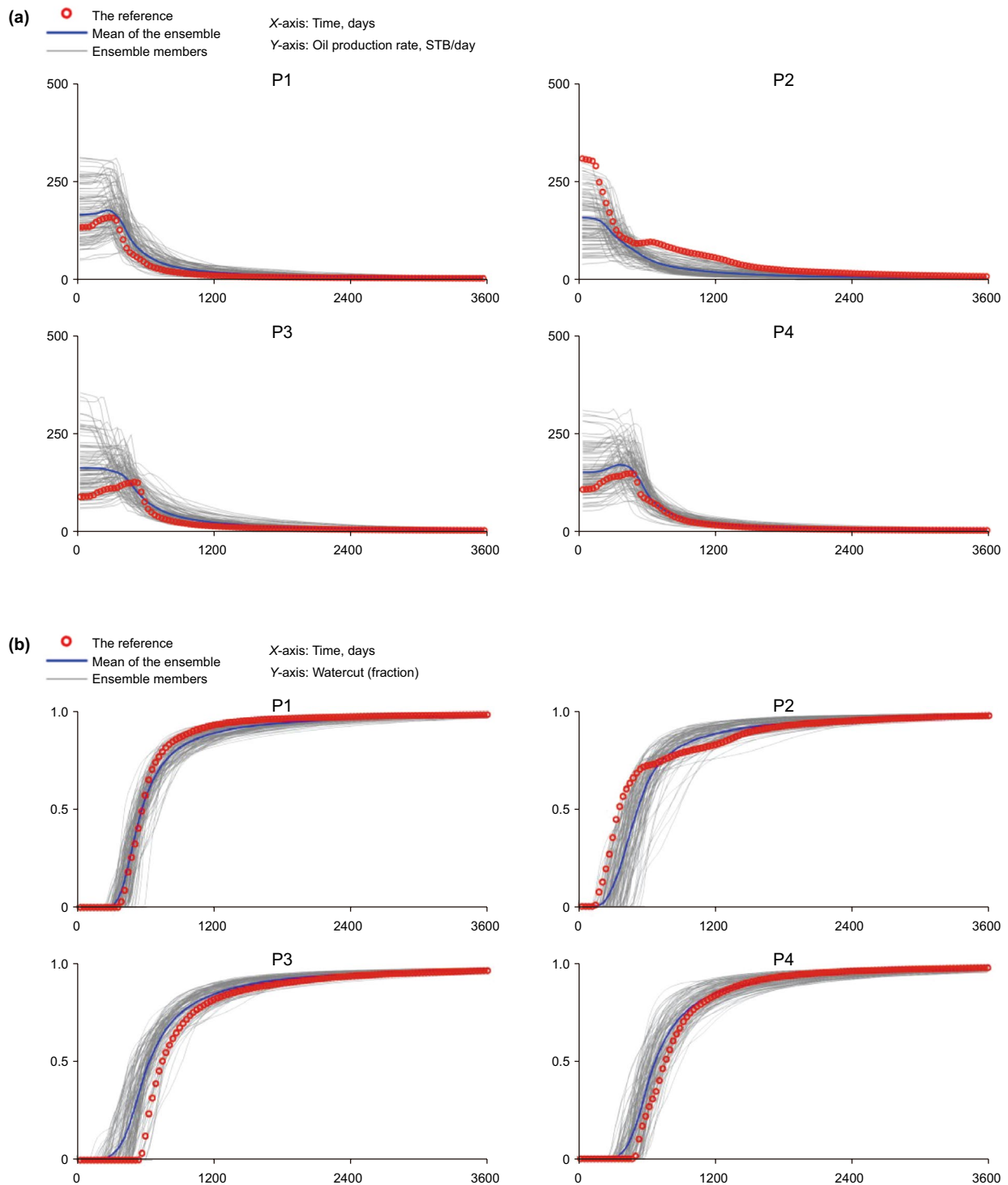

**Fig. 8** Egg model (revised from Jansen et al. 2014) (P: production well, I: injection well)

### 3.2 3D benchmark reservoir: egg model

We test our model selection scheme in another 3D channel reservoir model named the egg model (Jansen et al. 2014). It is one of the popular public benchmark models (Fig. 8). It consists of 101 ensemble members in total with a  $60 \times 60 \times 7$  grid system. It also has eight water injectors and four producers and we simulate for 3600 days. Jansen et al. (2014) assumed one of 101 ensemble members as the reference model, and the rest 100 models are used for history matching analysis.

Initial simulation results of the 100 ensemble members are compared on oil production rates and watercuts (Figs. 9 and 10). The early production stage before 1000 days has large uncertainty in the production wells. However, the mean value of the ensemble (dotted line) is similar to that of the reference (solid line) except for well P2.

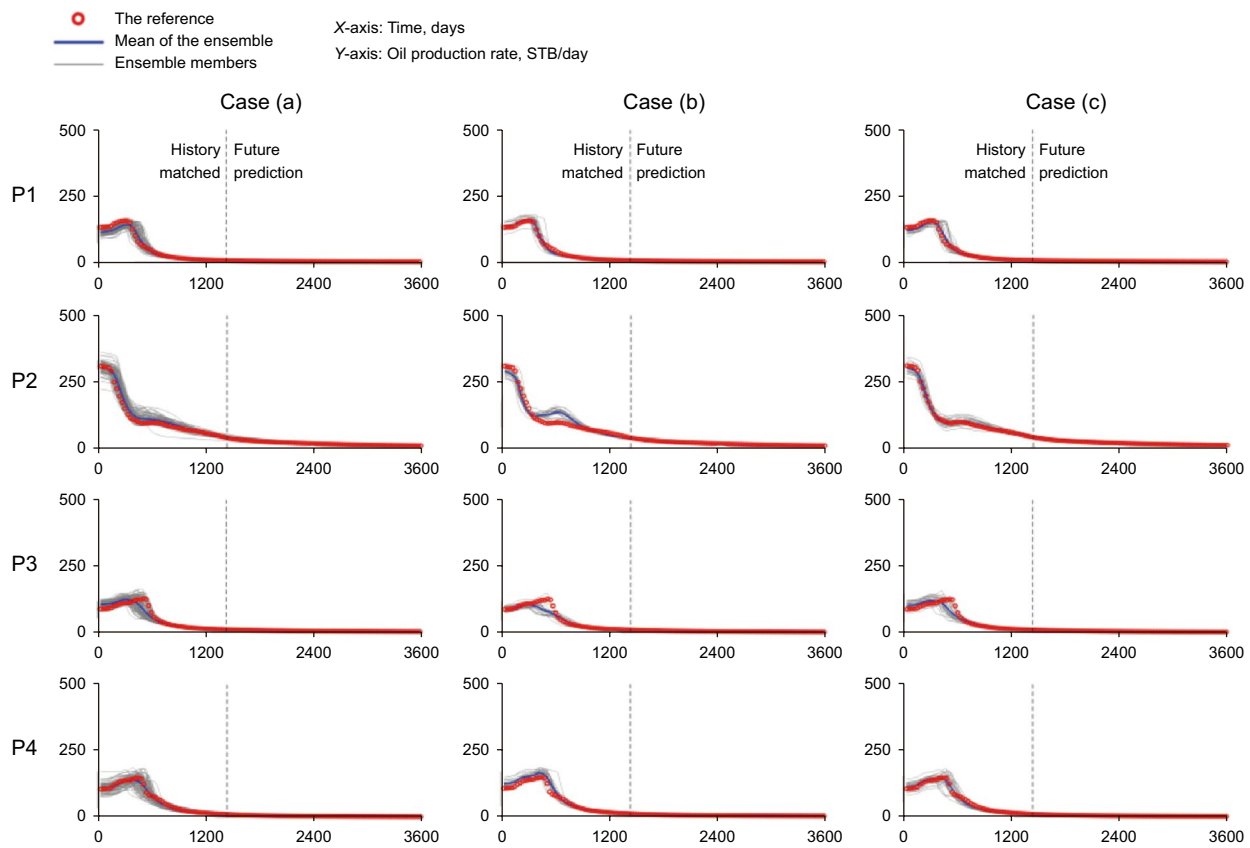
The oil production rates are observed every 360 days: 360th, 720th, 1,080th, and 1,440th days. The size of the initial ensemble is 100, and we select 30 reservoir models using the proposed PCA-assisted scheme. Three cases are compared: 100 models [case (a)]; 30 models by the uniform sampling method [case (b)]; 30 models by the proposed method [case (c)]. From Fig. 9, all ensemble members seem fairly similar with the reference so that we



**Fig. 9** Simulation results from the initial reservoir models on oil production rates and watercuts of the four producers in the egg model

may expect stable convergence when we use all the 100 models for EnKF. In the egg model case, we check again how the proposed selection scheme is efficient for fast and reliable analysis.

Oil and water productions of the three cases after EnKF updates are shown in Figs. 10 and 11, respectively. Overall trends of production forecasts seem to be reliable in all cases. However, note that the forecast ranges in case (b) are deviated from the reference, especially in well P2. Even with



**Fig. 10** Oil production rates of the four producers after EnKF in the three cases

the same ensemble size, case (c) gives reliable forecasts on the all producers. The results in case (c) can be comparable to the results using all the 100 models, but it takes only 33.1% of the simulation time, compared to case (a). It is the same trend when we draw the boxplots for the cumulative productions.

Tables 6 and 7 are results of quantitative analysis on production forecasts similar to the first case. Table 6 is the reduction ratio of prediction ranges of the production wells compared to case (a), and RMSEs are compared in Table 7. Since we only use 30% of well-made ensemble members in cases (b) and (c), the prediction ranges are reduced after the EnKF updates.

As can be seen in Figs. 10 and 11, the production forecasts from the realizations are similar. Therefore, quantitative analysis is needed to compare prediction accuracy among the three cases. RMSE is one of the good criteria to compare the results and as shown in Table 7, the proposed method gives the lowest RMSE in average.

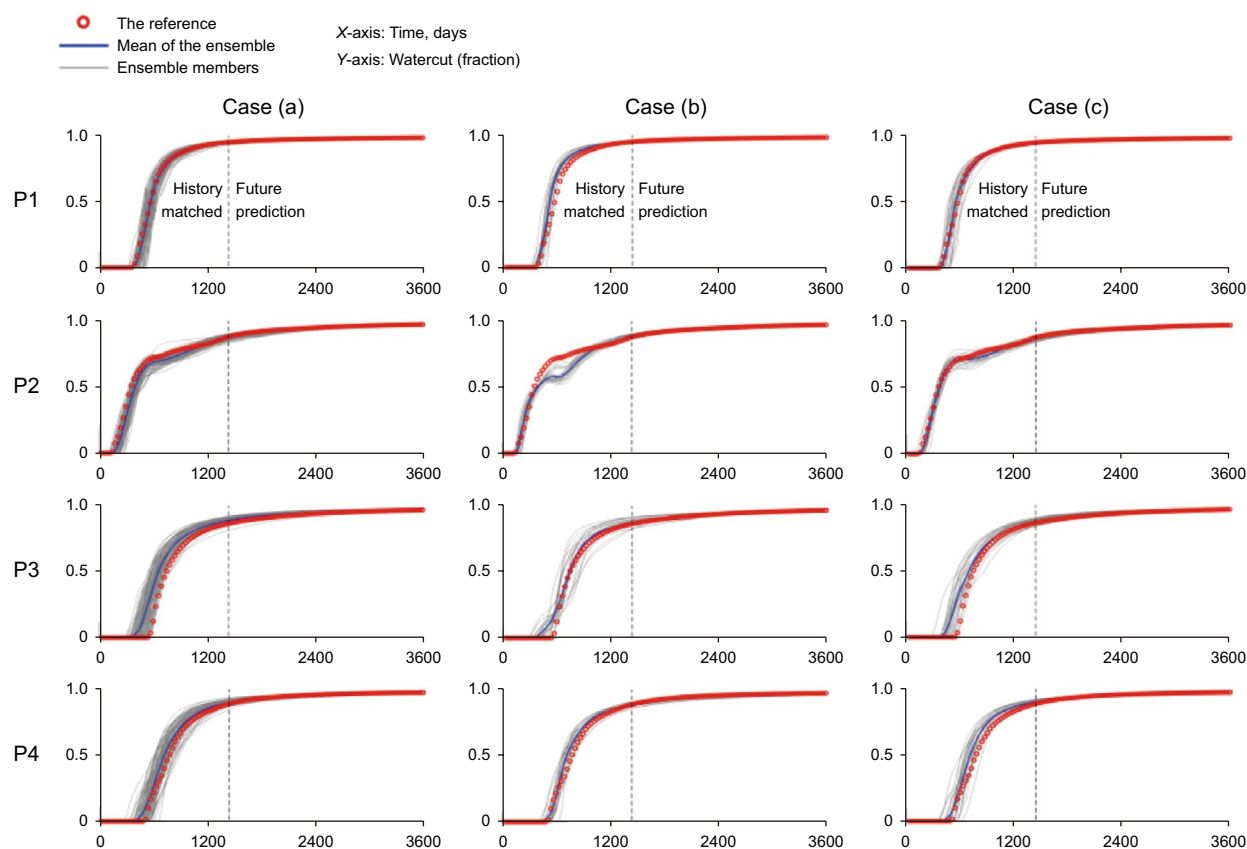
The normalized total oil and water productions at the end of the simulations are compared in boxplots (Fig. 12). Only the box in case (b) is out of the true line, which leads

to wrong production estimations. In case (c), we select 30 ensemble members from the 100 initial models. The future estimations in the boxplot in case (c) demonstrate that we can get more reliable estimations while using less computation times by using the proposed method.

## 4 Conclusions

Reservoir models are typically made by integrating static and dynamic data available. However, all data represent only some parts of the whole reservoir properties. Hence, uncertainty in reservoir models is natural and we propose one of the efficient strategies to manage the uncertain parameters for reliable analysis of future performances.

In this paper, the model selection scheme using PCA and clustering is proposed and applied to two complex 3D reservoir models. As a reservoir is a complex and large-scale grid system, selection of good reservoir models is critical for fast and reliable simulations. We study in a 3D channel reservoir and the egg model to demonstrate the effectiveness of the proposed sampling scheme. Compared to the case



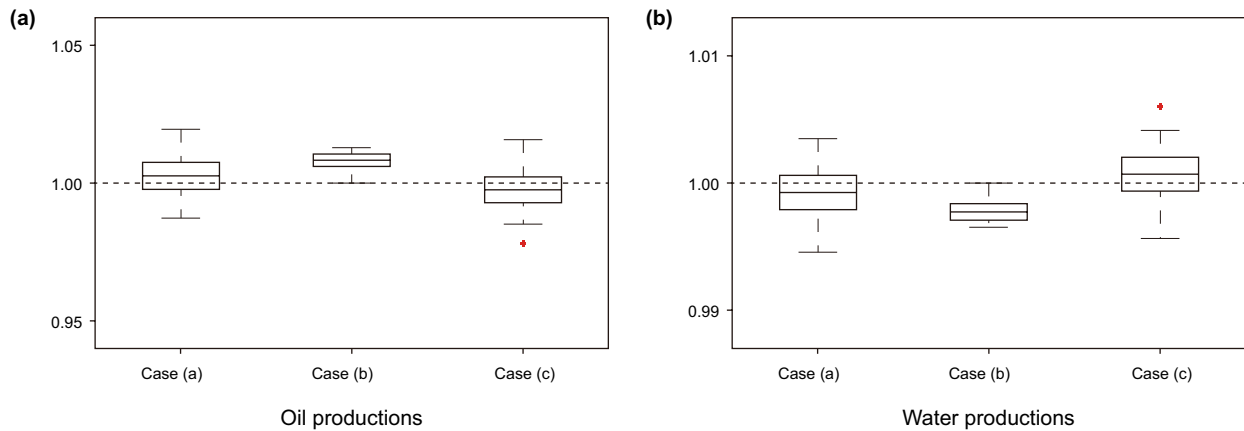
**Fig. 11** Watercuts of the four producers after EnKF in the three cases

**Table 6** Reduction ratio (%) of prediction ranges compared to case (a) in the egg model

Wells	Oil production rate			Watercuts		
	Case (a)	Case (b)	Case (c)	Case (a)	Case (b)	Case (c)
P1	0.000	12.679	34.170	0.000	29.314	29.055
P2	0.000	56.989	31.591	0.000	43.117	15.931
P3	0.000	9.316	−8.306	0.000	17.780	15.149
P4	0.000	59.390	48.927	0.000	31.836	36.514
Average	0.000	34.594	29.596	0.000	30.512	24.162

**Table 7** RMSE of oil production rates and watercuts in the egg model

Wells	Oil production rate			Watercuts		
	Case (a)	Case (b)	Case (c)	Case (a)	Case (b)	Case (c)
P1	0.878	1.494	0.331	0.003	0.018	0.003
P2	3.252	8.301	2.306	0.013	0.032	0.012
P3	0.943	0.928	1.000	0.043	0.012	0.027
P4	2.105	1.492	2.675	0.024	0.023	0.032
Average	1.794	3.054	1.578	0.021	0.021	0.018



**Fig. 12** Boxplots on the cumulative productions after 3600 days

with the same ensemble size, the results from the proposed method are more accurate on future production forecasts. Among the initial models generated by equiprobable geostatistical algorithms, there will be poor models that are quite different from the true reservoir model. From this study, we confirm that it is more important to have good initial models for stable assimilations by EnKF, rather than increasing the ensemble size.

The proposed method is to select good ensemble models prior to applying the optimization process such as EnKF. Therefore, it can be easily combined with other algorithms and give an effective solution for simulating 3D channel reservoirs in an efficient way.

**Acknowledgements** The authors thank research projects supported by The Ministry of Trade, Industry, and Energy (20172510102090, 20142520100440, 20162010201980), and Global PhD Fellowship Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2015H1A2A1030756). Hoonyoung Jeong has been supported by the National Research Foundation of Korea (NRF) Grant (No. 2018R1C1B5045260). The corresponding author is also thankful to The Research Institute of Engineering at Seoul National University, Korea.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

Arroyo-Negrete E, Devegowda D, Datta-Gupta A, Choe J. Streamline assisted ensemble Kalman filter for rapid and continuous reservoir model updating. *SPE Reserv Eval Eng*. 2008;11(6):1046–60. <https://doi.org/10.2118/104255-PA>.

- Chen C, Gao G, Ramirez B, Vink J, Girardi A. Assisted history matching of channelized models by use of pluri-principal-component analysis. *SPE J*. 2016;21(5):1793–812. <https://doi.org/10.2118/173192-PA>.
- Chen Y, Oliver DS. Cross-covariances and localization for EnKF in multiphase flow data assimilation. *Comput Geosci*. 2010;14(4):579–601. <https://doi.org/10.1007/s10596-009-9174-6>.
- Chiotoroiu MM, Peisker J, Clemens T, Thiele M. Forecasting incremental oil production of a polymer-pilot extension in the Matzen field including quantitative uncertainty assessment. *SPE Reserv Eval Eng*. 2017;20(4):894–905. <https://doi.org/10.2118/179546-PA>.
- Evensen G. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte Carlo methods to forecast error statistics. *J Geophys Res*. 1994;99(C5):10143–62. <https://doi.org/10.1029/94JC00572>.
- Evensen G. Sampling strategies and square root analysis schemes for the EnKF. *Ocean Dyn*. 2004;54(6):539–60. <https://doi.org/10.1007/s10236-004-0099-2>.
- Houtekamer PL, Mitchell HL. A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon Weather Rev*. 2001;129(1):123–37. [https://doi.org/10.1175/1520-0493\(2001\)129%3c0123:ASEKFF%3e2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129%3c0123:ASEKFF%3e2.0.CO;2).
- Huang Y, Zeng F. The temperature-based localization for the application of EnKF on automatic history matching of the SAGD process. *Comput Geosci*. 2016;20(1):187–212. <https://doi.org/10.1007/s10596-016-9557-4>.
- Jafarpour B, McLaughlin DB. Estimating channelized-reservoir permeabilities with the ensemble Kalman filter: the importance of ensemble design. *SPE J*. 2009;14(2):374–88. <https://doi.org/10.2118/108941-PA>.
- Jansen JD, Fonseca RM, Kahrobaei S, Siraj MM, Van Essen GM, Van den Hof PMJ. The egg model—a geological ensemble for reservoir simulation. *Geosci Data J*. 2014;1(2):192–5. <https://doi.org/10.1002/gdj3.21>.
- Jeong H, Ki S, Choe J. Reservoir characterization from insufficient static data using gradual deformation method with ensemble Kalman filter. *Energy Sources Part A*. 2010;32(10):942–51. <https://doi.org/10.1080/15567030802606350>.
- Jung H, Jo H, Kim S, Lee K, Choe J. Recursive update of channel information for reliable history matching of channel reservoirs using EnKF with DCT. *J Pet Sci Eng*. 2017a;154:19–37. <https://doi.org/10.1016/j.petrol.2017.04.016>.



- Jung H, Jo H, Lee K, Choe J. Characterization of various channel fields using an initial ensemble selection scheme and covariance localization. *J Energy Resour Technol Trans ASME*. 2017b;139(6):062906. <https://doi.org/10.1115/1.4037811>.
- Jung H, Jo H, Kim S, Lee K, Choe J. Geological model sampling using PCA-assisted support vector machine for reliable channel reservoir characterization. *J Pet Sci Eng*. 2018;167:396–405. <https://doi.org/10.1016/j.petrol.2018.04.017>.
- Kalman RE. A new approach to linear filtering and prediction problems. *J Basic Eng Trans ASME*. 1960;82:35–45. <https://doi.org/10.1115/1.3662552>.
- Kang B, Choe J. Regeneration of initial ensembles with facies analysis for efficient history matching. *J Energy Resour Technol Trans ASME*. 2017;139(4):042903. <https://doi.org/10.1115/1.4036382>.
- Kang B, Lee K, Choe J. Improvement of ensemble smoother with SVD-assisted sampling scheme. *J Pet Sci Eng*. 2016;141:114–24. <https://doi.org/10.1016/j.petrol.2016.01.015>.
- Kang B, Lee K, Choe J. Ensemble Kalman filter with principal component analysis assisted sampling for channelized reservoir characterization. *J Energy Resour Technol Trans ASME*. 2017;193(3):032907. <https://doi.org/10.1115/1.4035747>.
- Kim S, Jung H, Lee K, Choe J. Initial ensemble design scheme for effective characterization of three-dimensional channel gas reservoirs with an aquifer. *J Energy Resour Technol Trans ASME*. 2017;139(2):022911. <https://doi.org/10.1115/1.4035515>.
- Koneshloo M, Aryana SA, Grana D, Pierre JW. A workflow for static reservoir modeling guided by seismic data in a fluvial system. *Math Geosci*. 2017;49(8):995–1020. <https://doi.org/10.1007/s11004-017-9696-8>.
- Lee K, Jeong H, Jung SP, Choe J. Characterization of channelized reservoir using ensemble Kalman filter with clustered covariance. *Energy Explor Exploit*. 2013;31(1):17–29. <https://doi.org/10.1260/0144-5987.31.1.17>.
- Lee K, Jung SP, Choe J. Ensemble smoother with clustered covariance for 3D channelized reservoirs with geological uncertainty. *J Pet Sci Eng*. 2016;145:423–35. <https://doi.org/10.1016/j.petrol.2016.05.029>.
- Lee K, Jung SP, Lee T, Choe J. Use of clustered covariance and selective measurement data in ensemble smoother for three-dimensional reservoir characterization. *J Energy Resour Technol Trans ASME*. 2017;139(2):022905. <https://doi.org/10.1115/1.4034443>.
- Lim J, Jin J, Lee H, Choe J. Uncertainty analysis of oil sands reservoirs using features in metric space. *Energy Sources Part A*. 2015;37(16):1736–46. <https://doi.org/10.1080/15567036.2011.606457>.
- Liu N, Oliver DS. Ensemble Kalman filter for automatic history matching of geologic facies. *J Pet Sci Eng*. 2005;47(3):147–61. <https://doi.org/10.1016/j.petrol.2005.03.006>.
- Luo X, Bhakta T, Nævdal G. Correlation-based adaptive localization with applications to ensemble-based 4D-seismic history matching. *SPE J*. 2018;23(2):396–427. <https://doi.org/10.2118/185936-PA>.
- Luo X, Lorentzen RJ, Valestrand R, Evensen G. Correlation-based adaptive localization for ensemble-based history matching: applied to the Norne field case study. *SPE J*. 2019. <https://doi.org/10.2118/191305-PA>.
- Nævdal G, Mannseth T, Vefring EH. Near-well reservoir monitoring through ensemble Kalman filter. In: *SPE/DOE improved oil recovery symposium*. Society of Petroleum Engineers; 2002. SPE-75235-MS. <http://dx.doi.org/10.2118/75235-MS>.
- Nævdal G, Johnsen LM, Aanonsen SI, Vefring EH. Reservoir monitoring and continuous model updating using ensemble Kalman filter. *SPE J*. 2005;10(1):66–74. <https://doi.org/10.2118/84372-PA>.
- Patel RG, Trivedi J, Rahim S, Li Z. Initial sampling of ensemble for steam-assisted-gravity-drainage-reservoir history matching. *SPE J*. 2015;54(6):424–41. <https://doi.org/10.2118/178927-PA>.
- Remy N, Boucher A, Wu J. *Applied geostatistics with SGeMS: a user's guide*. Cambridge: Cambridge University Press; 2011.
- Sætrom J, Hove J, Skjervheim JA, Vabø JG. Improved uncertainty quantification in the ensemble Kalman filter using statistical model-selection techniques. *SPE J*. 2012;17(1):152–62. <https://doi.org/10.2118/145192-PA>.
- Scheidt C, Caers J. Uncertainty quantification in reservoir performance using distances and kernel methods—application to a west Africa deepwater turbidite reservoir. *SPE J*. 2009;14(4):680–92. <https://doi.org/10.2118/118740-PA>.
- Siena M, Guadagnini A, Della Rossa E, Lamberti A, Masserano F, Rotondi M. A novel enhanced-oil-recovery screening approach based on Bayesian clustering and principal-component analysis. *SPE J*. 2016;19(3):382–90. <https://doi.org/10.2118/174315-PA>.
- Suzuki S, Caumon G, Caers J. Dynamic data integration for structural modeling: model screening approach using a distance-based model parameterization. *Comput Geosci*. 2008;12(1):105–19. <https://doi.org/10.1007/s10596-007-9063-9>.
- van Leeuwen PJ, Evensen G. Data assimilation and inverse methods in terms of a probabilistic formulation. *Mon Weather Rev*. 1996;124:2898–913. [https://doi.org/10.1175/1520-0493\(1996\)124%3c2898:DAAIMI%3e2.0.CO;2](https://doi.org/10.1175/1520-0493(1996)124%3c2898:DAAIMI%3e2.0.CO;2).
- Vo H, Durlowsky L. A new differentiable parameterization based on principal component analysis for the low-dimensional representation of complex geological models. *Math Geosci*. 2014;46(7):775–813. <https://doi.org/10.1007/s11004-014-9541-2>.
- Vo H, Durlowsky L. Data assimilation and uncertainty assessment for complex geological models using a new pca-based parameterization. *Comput Geosci*. 2015;19(4):747–67. <https://doi.org/10.1007/s10596-015-9483-x>.
- Vo H, Durlowsky L. Regularized kernel PCA for the efficient parameterization of complex geological models. *J Comput Phys*. 2016;322:859–81. <https://doi.org/10.1016/j.jcp.2016.07.011>.
- Zhao Y, Reynolds A, Li G. Generating facies maps by assimilating production data and seismic data with the ensemble Kalman filter. In: *SPE symposium on improved oil recovery*. Society of Petroleum Engineers; 2008. SPE-113990-MS. <https://doi.org/10.2118/113990-MS>.