Summary

The properly carried out procedure of the history matching of a petroleum reservoir is crucial for the successful reservoir management. A petroleum reservoir based on Utica play geology developed under water flooding regime was automatically history matched with the Ensemble Kalman Filter (EnKF). The updated permeability field improves prediction of the reservoir performance, when time-dependent production data are incorporated into the model in addition to the time-invariant log data. The associated uncertainty was reported through reservoir model ensemble, and future reservoir performance was forecasted with a greater confidence due to updated permeability field. In addition, the Ensemble Kalman Smoother (EnKS) was studied with several modifications including normal score transformation of the model input parameters, localization of model parameter – observations covariance matrix, incorporation of measurement error into observations, and reduction of model parameters to further improve history matching.

Application of Ensemble Kalman Filter (EnKF) and Ensemble Kalman Smoother (EnKS) to a Synthesized Utica-based Petroleum Reservoir for Assisted History Matching

Yevgeniy Zagayevskiy, SPE, Hanzi Mao, Harsh Biren Vora, Hui Dong, Terry Wong, SPE, Dominic Camilleri, SPE, Courtney Beck, and Charles Wang, Landmark, Halliburton

The history matching of the petroleum reservoir, which is constructed primarily based on the static geologic data, is one of the main procedures to improve current petroleum reservoir model with time-dependent production data. The data assimilation techniques, The Ensemble Kalman Filter (EnKF) and Ensemble Kalman Smoother (EnKF) that work well with both static and dynamic model parameters, were proposed and investigated for assisted history matching petroleum reservoir characterization workflow. The numerical experiments conducted on a small 2D petroleum reservoir model flow simulated under waterflooding regime showed that a series of improvements are required to enhance history matching results. Both EnKF and EnKS performed well with proposed modifications for this 2D case when permeability was updated to history match instantaneous oil production rate at four production wells. Normal score transformation of the input model parameters values was proposed to avoid updated non-Gaussian distributions of the input parameters to drift toward Gaussian shape distributions, which are caused by the nature of the proposed history matching algorithms. The localization of the covariance matrix between model parameters and observations was required through weighting function kernel in order to avoid model ensemble collapse due to spurious covariance values. The introduction of the measurement error into production observations also helped keep model ensemble stable. For large

models, the grid block retaining function was applied to the model grid to select gridded input model parameters for the update based on their proximity to the production wells. By doing so, the number of the model parameters was reduced to ensure decrease in a computational time and ability of EnKF and EnKS to handle large models. The EnKF scenario with weighted covariance matrix and normal score transformation was applied to the synthesized 3D Utica shale play-based petroleum reservoir to improve quality of the underlying model through history matching. The history matching results were improved by incorporating both petrophysical permeability data and oil production rate data sampled in a span of fifteen years. The mismatch between predicted oil production rate and observed values was reduced, and uncertainty of the petroleum reservoir behavior was mitigated as more data were incorporated into the model. The improved petroleum reservoir model is deemed to produce more accurate forecast of the future reservoir performance with quantified uncertainty. By looking at the results from 2D and 3D models, the proposed enhanced methodology of data assimilation and history matching with the EnKF and EnKS was proven reliable. These proposed techniques have a potential to become a part of the standard history matching workflow for the oil and gas companies to perform petroleum reservoir model calibration that are developed under waterflooding and, perhaps, any other hydrocarbon extraction regimes.

Introduction

A successful management of the oil and gas project strongly depends on the underlying model of the petroleum reservoir. The reservoir model is characterized by static and dynamic parameters. Former ones don't vary in time and are represented for example by petrophysical and geomechanical properties, relative permeability curves, PVT curves, well locations, etc. On the other hand, the dynamic parameters vary in time. A few good examples of dynamic properties would be oil production rate, bottomhole pressure, gas/oil ratio, and 4D seismic attributes. These parameters are estimated based on the field measurements and surveys and are calibrated in time. Then, the static and dynamic data are assimilated into petroleum reservoir model to ensure its robustness for the accurate future prediction of the reservoir behavior.

The petroleum reservoir production history matching or simply history matching is a procedure that calibrates petroleum reservoir model to historical production data (Gilman and Ozgen 2013). It is an inverse problem that has multiple solutions. An illustrative example of history matching is shown in Fig. 1 through the instantaneous oil production rate at some fictitious well. The production data or observations, shown as blue dots, are measured on the field for the first four years. We can see that oil production rate increases up to year 3, where it reaches its maximum rate, and then starts decreasing afterwards. An initial guess reservoir model, shown as green line, is constructed based on primarily static parameter data such as petrophysical properties from the logging procedure. This initial model leads to a production rate, which underestimates the actual production rate. When the static parameters are

adjusted to take into account dynamic parameter data in the form of the production data, the adjusted petroleum reservoir model, shown as red line, matches the historical production data better than the initial reservoir model does. This calibration procedure of the petroleum reservoir model to production data is a history matching procedure. The static model parameters are adjusted in a way to ensure not only reproduction of static parameter data, e.g. porosity and permeability values or relative permeability curve, but also to reproduce dynamic data as closely as possible. It is deemed that the adjusted petroleum reservoir model produces better forecast of the future oil production rate, shown as red dashed line for years 4 to 6, compared to the initial reservoir model forecast, shown as green dashed line. Therefore, it is crucial to history match petroleum reservoir to achieve more accurate prediction of the reservoir performance by adjusting static parameters, on which dynamic parameters are based.

There are different techniques for history matching, both manual and assisted (or semi-automatic as there is still a need from the human input). All techniques strive to minimize mismatch between the predicted values from the reservoir model and observed data. This mismatch is usually formulated as an objective function of the optimization problem. The assisted history matching techniques are more automated in comparison to manual ones, and are very helpful when numerous parameters should be adjusted for history matching. The assisted history matching techniques can be broadly classified into two groups as gradient- based and non-gradient-based methods (Rwechungura et al. 2011). The gradient-based techniques such as steepest decent, Gauss-Newton, conjugate gradient, etc. require analytical understanding of the modeled system to establish sensitivity matrix, a matrix with partial derivatives of the dynamic parameters with respect to static parameters, from which objective function is derived. Therefore, these techniques are usually mathematically complex and require deep knowledge of the reservoir system under study, which is a tedious, time consuming process. The non-gradient-based methods are simpler to implement, don't require analytical understanding of the system under study, but usually more computationally intense due to their requirement to execute numerous number of the flow simulation runs. The examples of the non-gradient group members are evolutionary strategy, genetic algorithms, simulated annealing, mesh adaptive direct search, Bayesian optimization, etc.

Even though there is an abundance of the history matching techniques conventionally used in the oil and gas industry, most of them don't provide various scenarios of the reservoir performance and, therefore, associated uncertainty can be hardly quantified. In this paper, we propose to examine Ensemble Kalman Filter (EnKF) and Ensemble Kalman Smoother (EnKS), non-gradient-based assisted history matching techniques, to history match a realistic petroleum reservoir (Aanonsen et al. 2009; Evensen 2009). These algorithms are tested and demonstrated on a simple 2D waterflooding petroleum reservoir and on a synthesized 3D model shaped after modified real project in Utica shale play. These two methods utilize an ensemble or set of model realizations that equiprobably describe the petroleum

reservoir under investigation. This allows us to predict uncertainty associated with the history matched petroleum reservoir model. A petroleum reservoir model constructed based on the Utica shale play represent a realistic reservoir and acts as a plausible reservoir to examine EnKF and EnKS and resolve major challenges associated with their implementation to real projects. On a downside, these algorithms are looked upon as computationally expensive ones. In order for its implementation to be practical, EnKF/EnKS should be applied in a distributed computing environment (with cloud technology) or at least be parallelized on a local machine, if full physics flow simulation runs are performed. Also, commercial flow simulators can be replaced with proxy flow models to additionally reduce computational time (Mishra and Datta-Gupta 2017).

There is a great number of works on practical application of the algorithms to history matching petroleum reservoirs. Evensen et al. (2007) and Haugen et al. (2008) applied EnKF to history match a North Sea petroleum reservoir. Zhang and Oliver (2011) applied EnKF to history match production from deep water reservoir PFJ2 in the Gulf of Mexico region. Patel et al. (2015) compared EnKF to other history matching techniques, such as *K*-means clustering, importance sampling, and sampling based on the orthogonal ensemble members, on steam assisted gravity drainage (SAGD) case study of a northern Alberta oil sands field. De Lima et al. (2017) tested EnKF-based history matching on a large heavy-oil turbidites reservoir, one of the offshore fields in Campos Basin of Brazilian coast. Lin et al. (2017) applied EnKS to history match a large field-scale reservoir in the Middle East region of 60 million grid blocks. Ma et al. (2017) also used EnKS to history match two synthesized highly nonlinear petroleum reservoir models. In all these and other cases results were successful. This proposed paper is different in a way that we are comparing EnKF and EnKS, proposing a series of improvements, and applying EnKF to modified Utica data set.

This paper is organized in the following manner. First, history matching workflows based on the EnKF and EnKS are provided and theoretical details are explained as a solution to the petroleum reservoir model calibration. Second, the implementation improvements tailored to Utica data set are proposed and demonstrated on a smaller 2D case study. EnKF and EnKS algorithm performances are compared as well. Third, the Utica shale play is presented and reviewed. Fourth, the EnKF workflow is applied to fine tune synthesized Utica-based petroleum reservoir model to history match production data. The results are shown and discussed. Lastly, the conclusions are made, and future work is outlined.

Description of EnKF and EnKS-based History Matching

The methodology of the EnKF and EnKS-based history matching for the petroleum reservoirs and theoretical details of the EnKF forecasting and updating steps are presented in this section.

Workflow of the Proposed History Matching Algorithms. The workflow of history matching is shown in Fig. 2 for EnKF and in Fig. 3 for EnKS. In the EnKF, the main idea is to continuously update static model parameters, for example gridded permeability, by incorporating production data sequentially from one time step at a time. The possible production data are oil production rate, water production rate at production well locations, water injection rate at injection well locations, and much less time varying bottomhole pressure at all wells. The updated permeability should allow better prediction of the production data by the petroleum reservoir model. Other petrophysical parameters such as porosity model or other parameters as relative permeability curves can be potentially updated with the exactly same history matching methodology based on the EnKF.

In contrast to EnKF, in EnKS all available data are assimilated at once (Fig. 3). Because a petroleum reservoir model is nonlinear system, an iterative assimilation of the data is performed for EnKS to improve relationship between model input parameter values and corresponding output parameter values after the update.

In this paper we operate only with the following reservoir model parameters: static parameter permeability (KX) $k(n)$, which is updated at each time step t_n with n being the iteration number, $n = 0, \ldots, N$, and is resolved on a 2D/3D grid; dynamic parameters – oil production rate (OPR) $q_{oil}(t_{0:n})$ for 2D example, and additionally water production rate (WPR) $q_{water}(t_{0:n})$ for 3D example resolved at five production well locations. However, there is no need to limit another study only to the aforementioned parameters; data from other sources can be potentially incorporated into the reservoir model and, therefore, corresponding model parameters should be considered. The log data are used to build initial permeability ensemble. The production data, which are field measurements of $q_{oil}(t_n)$ and $q_{water}(t_n)$, are incorporated at each time step t_n.

We will review EnKF in more detail by following diagram in Fig. 2, as EnKS is very similar to EnKF. Our main input parameter that will be changed during the history matching is the X component of horizontal permeability KX. We assume that the Y component of horizontal permeability equals to KX and vertical component KZ is a tenth of KX. In this text, we will refer to KX as permeability for the sake of simplicity. In the very beginning of history matching with EnKF workflow, an initial ensemble of N_R permeability realizations is simulated with sequential Gaussian simulation (SGS) geostatistical algorithm conditional to permeability data at well locations derived from log measurements. These permeability realizations are used as input parameters to a flow simulator, which runs N_R flow simulation from previous time step t_0 to a current time step t_1 to produce an ensemble of predicted model response values such as production rate at the production well locations. Then, the covariance matrix between permeability values in the model grid and production variables at well locations is computed from the corresponding realizations. This covariance matrix is used to derive coefficients that are needed to update permeability field while assimilating the production data. Once the permeability model is updated, the

process is repeated for the next time step t_2 as shown in diagram of Fig. 2. Once all production data are assimilated into the model at the time step t_n, the final updated permeability field is used to make a forecast of the production parameters such as oil production rate in the future (time t_{n+1}, t_{n+2}, …). The resulting ensemble of production rates can be used to assess uncertainty in the prediction by observing the variation in the ensemble. It is deemed that final updated permeability model is brought closer to the true permeability field by calibrating the entire petroleum reservoir model to the available production data from all time steps. Note that the modeling of the lithological facies has been omitted in the proposed approach. Only continuous properties are explored and categorical properties are left behind for now (Pyrcz and Deutsch 2015). The proposed general EnKF-based workflow can be applied to any petroleum reservoir model for history matching.

As was mentioned before, in EnKS all available data from time steps t_1, …, t_n are assimilated simultaneously with the help of covariance matrix between updated model parameters, for example values of permeability in certain grid blocks, and assimilated observed production data, such as oil production rate from production wells. This covariance matrix is computed from the ensemble members. The proposed difference between EnKS and EnKF is that here the data assimilation step for all available production data happens multiple times until a target mismatch between flow simulation predictions and observation data is reached.

Theoretical Background of the EnKF and EnKS. Fig. 4 presents updating step of the EnKF in more detail. The entire EnKF procedure for a single assimilation step n can be summarized as follows through two step procedure: model forecast and model update.

In the model forecast step, the previously updated permeability ensemble $k(n\text{-}1)$ is used as an input to run a flow simulation for each realization. Let's denote this ensemble of static parameters as matrix $\mathbf{M}_1(n\text{-}1)$ of N_{M1} x N_R size. Here, N_{M1} is the number of active grid blocks with permeability values in the reservoir model, and n-1 is the iteration index. The forecast step of the EnKF is the flow simulation step in the proposed workflow, and is not emphasized as being part of the EnKF procedure explicitly.

The forecast step of the EnKF can be expressed analytically as shown in Eq. (3). Here, f denotes a flow simulation operator. We collectively address this ensemble of dynamic variables as matrix $\mathbf{M}_2(t_n, n\text{-}1)$ of size N_{M2} x N_R, where N_{m2} is the number of dynamic variables of the petroleum reservoir model.

$$\mathbf{M}_2(t_n, n-1) = f(\mathbf{M}_1(n-1), \mathbf{M}_2(t_{n-1}, n-1)) \quad \text{(3)}$$

Once the flow simulation is performed for each ensemble member, $q_{oil}(t_n)$ and $q_{water}(t_n)$ become available at the well locations. Note that dynamic parameters $\mathbf{M}_2$ change from time step t_{n-1} to t_n

according to some physical laws like mass balance and Darcy flow, while static parameters $\mathbf{M}_1$ are only updated numerically to newly assimilated data and don't physically change in time.

The augmented matrix $\mathbf{M}(t_n, n\text{-}1)$ contains both static $\mathbf{M}_1(n\text{-}1)$ and dynamic $\mathbf{M}_2(t_n, n\text{-}1)$ parameters; and its size is N_M x N_R, where $N_M = N_{M1} + N_{M2}$. Therefore, the covariance matrix $\mathbf{C}(n\text{-}1)$ of the entire petroleum reservoir model has size of N_M x N_M and is computed as shown in Eq. (4). Here, **u** is the N_R x 1 column vector of ones, and ^{T} is the matrix transpose operator. We will need this covariance matrix (or to be precise, just part of it that contains model parameter – observations covariance values) in the model update step.

$$\mathbf{C}(n-1) = \frac{1}{N_R - 1}\left(\mathbf{M}(t_n, n-1) - \frac{1}{N_R}\mathbf{M}(t_n, n-1)\mathbf{u}\mathbf{u}^T\right)\left(\mathbf{M}(t_n, n-1) - \frac{1}{N_R}\mathbf{M}(t_n, n-1)\mathbf{u}\mathbf{u}^T\right)^T \quad ..(4)$$

In the model update step, the entire petroleum reservoir model matrix $\mathbf{M}(t_n, n)$ is updated according to Eq. (5). The EnKF update step is linear and tends to enforce distribution of the modified parameters to be Gaussian. Because of linear updating nature of the EnKF, the physical relationship between static model parameters $\mathbf{M}_1(n)$ and dynamic variables $\mathbf{M}_2(t_n, n)$ most likely doesn't hold any longer and should be updated by rerunning flow simulation for the same time steps from $t_{n\text{-}1}$ to t_n (which is omitted in this paper for the sake of the computational time efficiency). Here, the field data is stored in a vector $\mathbf{d}(t_n)$ of size $N_{d(tn)}$ x 1; $\mathbf{D}(t_n)$ is the data matrix of size $N_{d(tn)}$ x N_R that is generated from the observed data by adding randomly distributed noise with zero mean and diagonal data covariance matrix $\mathbf{C}_d(t_n)$ estimated from the data measurement error; $\mathbf{H}(t_n)$ is the observation matrix of size $N_{d(tn)}$ x N_M, where each row consists of zero values except one entry, which equals to 1 and corresponds to the location in the observation matrix of the same order as a model parameter being measured in the model matrix $\mathbf{M}(t_n, n\text{-}1)$; and lastly, $\mathbf{K}(n\text{-}1)$ is the Kalman gain matrix of size N_M x $N_{d(tn)}$ that contains updating weights of the EnKF. Note that we are primarily interested in $\mathbf{M}_1(n)$ part of the entire model matrix $\mathbf{M}(t_n, n)$, where updated permeability values $k(n)$ are stored.

$$\mathbf{M}(t_n, n) = \mathbf{M}(t_n, n-1) + \mathbf{K}(n-1)\cdot\left(\mathbf{D}(t_n) - \mathbf{H}(t_n)\cdot\mathbf{M}(t_n, n-1)\right) \quad \text{.....(5)}$$

The Kalman gain is computed as shown in Eq. (6). Here, $^{-1}$ is the matrix inverse operator.

$$\mathbf{K}(n-1) = \mathbf{C}(n-1)\cdot\left(\mathbf{H}(t_n)\right)^T\cdot\left(\mathbf{H}(t_n)\cdot\mathbf{C}(n-1)\cdot\left(\mathbf{H}(t_n)\right)^T + \mathbf{C}_d(t_n)\right)^{-1} \quad \text{.....(6)}$$

After the ensemble updating is performed, the EnKF two-step procedure is repeated for the next assimilation step n+1, where updated permeability values in $\mathbf{M}_1(n)$ are used in the flow simulation to predict production parameters of the reservoir model for the next time step t_{n+1}.

The forecast and update steps of the EnKS can be expressed as shown in Eqs. (7) and (8), where $\mathbf{M}(n\text{-}1)$ contains $\mathbf{M}_1(n\text{-}1)$ and $\mathbf{M}_2(t_{0:Nt}, n\text{-}1)$ at the iteration step n-1 with $\mathbf{M}_2$ considering all time steps $t_{0:Nt}$ (from t_0 to t_{Nt}). N_t is the total number of the time steps.

$$\mathbf{M}_2(t_{0:N_t}, n-1) = f(\mathbf{M}_1(n-1), \mathbf{M}_2(t_0, n-1)) \quad \text{.....(7)}$$

$$\mathbf{M}(n) = \mathbf{M}(n-1) + \mathbf{K}(n-1) \cdot \left(\mathbf{D}(t_{0:n}) - \mathbf{H}(t_{0:n}) \cdot \mathbf{M}(n-1)\right) \quad \text{.....(8)}$$

Improvements to EnKF and EnKS Algorithms

A number of the challenges are foreseen for the implementation of the EnKF-based history matching workflow to a history matching problem of a real reservoir scale. First of all, the history matching results should show a significant improvements of the initial reservoir model through assimilation of dynamic data with possibility of uncertainty quantification in the updated model. Other crucial points to address are to ensure that chosen number of realizations is big enough to compute stable covariance matrix and small enough to avoid high computational cost of the flow simulation runs and EnKF update, to choose right data for the integration into the model, and to handle parameters with highly non-Gaussian distributions, e.g. skewed distribution. The following improvements to the EnKF and EnKS algorithm are suggested to overcome all aforementioned challenges. All improvements are first demonstrated on a simple 2D heterogeneous case study, which is similar to Utica problem from geological and petroleum engineering point of view, and then some of the improvements to the algorithms are applied to Utica case study.

The 2D heterogeneous examples are generated on N_X = 51 by N_Y = 51 grid (thus, there are 2,601 grid blocks in total) with single injection well placed at the center of the model in cell (26, 26), and four production wells are placed at the corners of the model at cells (2, 2), (50, 2), (50, 50), and (2, 50) as shown in Fig. 5. Each cell has a size of 10.0 x 10.0 square feet. The bottom-hole pressure was fixed at 10,000 psia for injection well and at 4,800 psia for all production wells. The production continues for 9 years with one year time step. The true reservoir representation (base case) is known and all constructed models are compared to it. The available data are permeability values at all five well locations and oil production rate at production well locations for the span of 9 years. The ensemble model consists of N_R = 50 realizations. The normal score for permeability fields, both true and initial ensemble members, were generated with SGS geostatistical algorithm. The variogram model to simulate the normal scores of permeability fields is exactly the same as for true case and ensemble members, form of which is

presented in Eq. (9) (Deutsch and Journel 1998). This is a spherical isotropic variogram model with variogram range *a* = 150.0 feet.

$$\gamma(h) = \begin{cases} 1.5h/a - 0.5(h/a)^3, & if\ h \le a \\ 1.0, & if\ h > a \end{cases} \quad \text{.....(9)}$$

where, *h* is the lag distance between two locations in space; and *a* is the variogram range.

The permeability values in mD units are computed from the normal score values as shown in Eq. (10) to ensure that permeability has a skewed distribution. The choice of this equation is arbitrary.

$$k = 50e^{0.75u} \quad \text{.....(10)}$$

where, *k* is the permeability values in units of mD; and *u* is the normal score values generated with SGS that follow Gaussian distribution with zero mean and unit variance.

The resulting maps of true permeability field and all ensemble members are shown in Fig. 5 with histogram plots of these permeability fields being presented in Fig. 6. The corresponding production characteristics in the form of instantaneous oil production rate for 4 production wells are plotted in Fig. 7.

Eleven different cases are examined on 2D reservoir model and compared to each other as summarized in Table 1. Nexus flow simulator was used to predict pressure and saturation distributions in space and time along with well rates. First two cases utilize EnKF, where original permeability values are updated in case 1, and permeability values are transformed to Gaussian distribution with normal score transformation (NST) procedure in case 2 (Pyrcz and Deutsch 2014). Remaining 7 cases deal with EnKS. Case 3 and 4 are similar to case 1 and 2 with difference of EnKF being replaced with EnKS. Cases 5, 6, and 7 are implemented with covariance localization function, which is based on a distance function, of various effective range. The assimilated data are assumed to be error free for all cases except cases 8 and 9 of various error magnitude. Cases 10 and 11 are prepared to show how number of updated input model parameters can be reduced through introduction of a grid block retaining function. Case 11 combines grid block retaining function with covariance localization function to minimize update artifact in the permeability field.

Table 1 – List of eleven cases to learn effect of algorithm parameters on history matching results

Case #	Algorithm	Data Assimilation	Normal Score Transfor-mation	Weighted Model Parameters – Observations Covariance Matrix	Data Error	Model Size Reduction	Average RMSE at last iteration,	Average RMSE at last iteration,

							KX	OPR
0	Initial ensemble	-	-	-	-	-	7631.51	513494
1	EnKF	Sequentially	No	No	No	No	6349.15	500024
2	EnKF	Sequentially	Yes	No	No	No	5790.14	207377
3	EnKS	All at once	No	No	No	No	6627.59	116890
4	EnKS	All at once	Yes	No	No	No	13452.20	21423
5	EnKS	All at once	No	Yes, long radius	No	No	5339.51	116172
6	EnKS	All at once	No	Yes, medium radius	No	No	4693.67	196817
7	EnKS	All at once	No	Yes, short radius	No	No	5518.34	326767
8	EnKS	All at once	No	No	Yes, large	No	4806.51	84176
9	EnKS	All at once	No	No	Yes, small	No	4956.36	81070
10	EnKS	All at once	No	No	No	Yes	6661.17	408171
11	EnKS	All at once	No	Yes, medium radius	No	Yes	5122.49	241022

The details of each improvement will be covered in more depth below.

The history matching metrics are provided in Table 1 for permeability field and oil production rate at 4 wells through average root mean square error (<RMSE>) for last iteration $n = N$ as presented in Eq. (11) and (12), respectively. Note that for cases 1 and 2 $N = 10$, while for cases 3 – 11 $N = 5$. The individual RMSE values for permeability and oil production rate are computed as shown in Eq. (13) and (14).

$$\left\langle RMSE_k(N)\right\rangle = \frac{1}{N_R}\sum_{r=1}^{N_R} RMSE_{k^r}(N) \quad \text{.....(11)}$$

$$\left\langle RMSE_{q_{oil}}(N)\right\rangle = \frac{1}{N_R}\sum_{r=1}^{N_R} RMSE_{q_{oil}^r}(N) \quad \text{.....(12)}$$

$$RMSE_{k^r}(n) = \frac{1}{N_X N_Y}\sqrt{\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y}\left(k_{i,j}^r(n) - k_{i,j}^{true}\right)^2}, \quad r = 1,\ldots,N_R, \quad n = 1,\ldots,N \quad \text{.....(13)}$$

$$RMSE_{q_{oil}^r}(n) = \frac{1}{N_w N_t}\sqrt{\sum_{i=1}^{N_w}\sum_{j=1}^{N_t}\left(q_{oil,i,j}^r(n) - q_{oil,i,j}^{true}(n)\right)^2}, \quad r = 1,\ldots,N_R, \quad n = 1,\ldots,N \quad \text{..(14)}$$

where, N is the total number of the iterations; r is the realization index; $k^r_{i,j}$ is the simulated permeability value of r^{th} realization in grid block (i, j); $k^{true}_{i,j}$ is the true permeability value in grid block (i, j);

$q^{r}_{oil,i,j}$ is the simulated annual instantaneous oil production rate value of r^{th} realization at i^{th} well in j^{th} year; and $q^{true}_{oil,i,j}$ is the true annual instantaneous oil production rate value of r^{th} realization at i^{th} well in j^{th} year.

Fig. 5 – 6 show how permeability fields were updated for all 11 cases for first realization. Fig. 8 – 12 present history matching plots for oil production rates at 4 wells for all 11 cases through iterations. Fig. 13 summarizes average RMSE. Fig. 14 shows how RMSE improved with iterations for permeability ensemble. Fig. 15 shows similar metrics for oil production rate. Fig. 16 shows weighting function for all 11 cases. They will be explained in the following sections.

Based on these figures and RMSE values from Table 1, both algorithms improve history matching results, but in general EnKS algorithm performs better than EnKF. Also, normal score transformation of the input variable prior to the update is suggested to improve accuracy of the history matching results by preserving shape of original distribution of the updated variable (in these cases it is permeability field). In this case, we would avoid seeing for example negative values of the permeability in the updated fields. For these cases, we enforced all negative values of the permeability to be zero for the flow simulator to work. Either covariance matrix localization or consideration of the error in the assimilated data is suggested to avoid ensemble collapse, when all realizations become exactly the same or very close to each other. The weighting function for covariance matrix localization with longer effective radius (when original covariance matrix does not change much), leads to a better history matching results in comparison to weighting functions with shorter radii. Consideration of error in the data not only enforces avoidance of the ensemble collapse, but also improves history matching results. Localization of the covariance matrix is required, when the size of the model is reduced through cookie-cutter type grid block retaining function, not only to avoid geologically unrealistic artifacts in the updated permeability field, but also to improve history matching results. The general conclusion from analysis of permeability and oil production rate convergence charts is that both more data and more iterations improve history matching results and concurrently minimize uncertainty.

Normal Score Transformation to Preserve Original Distribution Shape. One of the main improvements to the conventional implementation of the EnKF or EnKS is the application of the normal score transformation procedure (NST) to a permeability field values, in case they do not follow Gaussian distribution. This transformation is required, since the updated permeability field with the EnKF tends to be more Gaussian than should be, when the underlying true distribution is far from being Gaussian.

The NST procedure is well documented technique to transform any data set of arbitrary distribution to a normal distribution by matching quantiles of the cumulative probability distribution functions in original units and normal scores (Deutsch and Journel 1998). The NST procedure and reverse NST procedure can be analytically expressed as in Eq. (15) and Eq. (16). Here, $s(n)$ is the normal score value that corresponds to permeability $k(n)$ value in original units (most of the time in units of mD) at the

n^{th} iteration; $F(k(n))$ is the cumulative probability distribution function of permeability; $G(s(n))$ is the cumulative probability normal distribution function of normal scores; and $^{-1}$ is the inverse of a function. The EnKF and EnKS forecast step uses permeability in original units. The update step requires NST for the entire permeability ensemble. Once the normal scores are updated to newly assimilated production data, the normal scores values are converted back to the original units of permeability field using a look up table.

$$s(n) = G^{-1}\left[F\left(k(n)\right)\right] \quad \text{(15)}$$

$$k(n) = F^{-1}\left[G\left(s(n)\right)\right] \quad \text{(16)}$$

Case 1 versus case 2 and case 3 versus case 4 clearly show that NST procedure applied to permeability field prior to the update step and inverse of NST procedure applied to permeability field after it has been updated in normal score units improve accuracy of the petroleum reservoir model significantly. This claim can be seen numerically in Table 1 through average RMSE value and graphically in Fig. 5, 6, 8 – 11. Therefore, it is recommended to update values of the input parameters that follow highly non-Gaussian distribution in normal scores.

Covariance Matrix Localization for Ensemble Stability. For more accurate EnKF and EnKS update of the model, an ensemble should consist of infinitely many realizations. In practice, we usually limit ourselves to an ensemble of less than 100 realizations to ensure feasible running time of the flow simulator for complex petroleum reservoirs. However, depending on the number of the assimilated data and number of the parameters in the model, there is a chance that computed covariance matrix is not stable and parameters that should not be spatially correlated appear to be correlated. These covariance values are spurious due to the selected limited sample size that possibly does not accurately represent the entire parameter space. We propose to apply additional weights to the gridded variables, such as permeability, in updating part of the EnKF to mitigate the presence of spuriousness in the covariance matrix. As we saw from cases 5 – 7, the weighting function helps avoid the ensemble collapse and improves permeability field prediction.

The proposed equation to compute additional weights for localization of the updated model parameters – observed variables covariance matrix are based on the distance function that measure the minimum distance from given grid cell to the well location. The equation of the weights ω for an arbitrary cell (i, j) is calculated as shown in Eq. (17). Here, $b(i, j)$ is the distance function value assigned to a cell (i, j) on the grid, which is calculated according to Eq. (18); min is the operator that chooses the minimum distance function value among all distance function values calculated for a grid cell relative to all wells; N_X

and N_Y are the number of the cells in a regular grid in *X* and *Y* directions, respectively; and α is the power that allows to control the subsidence rate of the weights as the distance from the wells increases. The weights are one at all well locations and decrease to zero as a cell lies further and further from any well location: $\omega(i, j) \in [0,1], \forall i, j$. Maps of weights for cases 5 – 7 are shown in Fig. 16, where α = 0.3, α = 1.0, and α = 3.0 for case 5, 6, and 7, respectively.

$$\omega(i, j) = \left(1.0 - \frac{b(i, j)}{\min(b, \forall i, j)}\right)^{\alpha}, \quad i = 1,...,N_X, \quad j = 1,...,N_Y \qquad (17)$$

In Eq. (18) for distance function value $b(i, j)$ calculation, *w* is the index of the wells on the grid; *x* and *y* are the coordinates of either centers of grid cells or locations of the production wells on *X* and *Y* axes.

$$b(i, j) = \min\left(\sqrt{\left(x(i, j) - x_w\right)^2 + \left(y(i, j) - y_w\right)^2}, \forall w\right) \qquad (18)$$

Once weights are calculated for all cells in the grid, the weighting matrix **ω** is formed of size N_M x N_R. Here, each column consists of $\omega(i, j)$ values computed according to Eq. (17) that are assigned to corresponding static parameter stored in the model matrix **M**. Thus, the first column repeats each other for the rest N_R – 1 columns of additional weighting matrix **ω**. The modified updating equation of the EnKF is presented in Eq. (19). Here, ○ is the Hadamard or entrywise product.

$$\mathbf{M}(n) = \mathbf{M}(n-1) + \boldsymbol{\omega} \circ \mathbf{K}(n) \cdot \left(\mathbf{D}(t_n) - \mathbf{H}(t_n) \cdot \mathbf{M}(n-1)\right) \qquad (19)$$

The weights can be further explored by adjusting power α in Eq. (17) or by replacing distance function in Eq. (18) with a different form, e.g. anisotropic form, or the from correlated to the water flooding front.

Data Error for Ensemble Stability. In another approach to covariance matrix stabilization, the data error should not be ignored as it was done for most cases except for cases 8 and 9. When $\mathbf{C}_d(t_n) = 0.0$ for all t_n Kalman gain equation Eq. (6) takes a form shown in Eq. (20).

$$\mathbf{K}(n-1) = \mathbf{C}(n-1) \cdot \left(\mathbf{H}(t_n)\right)^T \cdot \left(\mathbf{H}(t_n) \cdot \mathbf{C}(n-1) \cdot \left(\mathbf{H}(t_n)\right)^T\right)^{-1} \qquad (20)$$

But, if $\mathbf{C}_d(t_n) \neq 0.0$, then history matching and permeability updating results are slightly improved as the covariance matrix is stabilized. We can see this in Table 1 and Fig. 5, 6, 10, 13, 14, and 15 when cases 8 and 9 are compared to case 3 for EnKS. All realizations do not collapse into a single one when the data error is not ignored. Therefore, it is recommended not to completely ignore error in the assimilated production data.

Grid Block Retaining Function for Model Parameter Reduction. In case when the model is too big, and it is impractically to compute entire covariance matrix of the model, it is suggested to work with reduced number of the variables. The widely used approach is the parameter reduction techniques like principal component analysis (PCA) or discrete cosine transform (DCT) (Rwechungura et al. 2011). Instead, we will limit the updating part only to the certain area around the wells by operating only on a limited number of the grid blocks. It is advised to apply this approach along with the previously discussed weighting function to avoid sharp boundaries in the permeability model after an update as shown in cases 10 and 11.

The local updating approach requires reducing number of the grid blocks by applying a 'cookie-cutter' function φ to the all grid blocks, which is based on the weighting function ω and a threshold $\omega^{threshold}$ specified by a user. Analytically it can be expressed as shown in Eq. (21). If the weight $\omega(i, j)$ is higher than threshold value, the grid block retaining function $\varphi(i, j)$ at this grid cell is one, otherwise it takes zero value. The grid block retaining function can change over time to include more grid blocks further away from the well locations and omit the blocks adjacent to the wells, as the petroleum reservoir model becomes more knowledgeable with time about reservoir properties at short distances away from the wells.

$$\varphi(i, j) = \begin{cases} 1, & if\ \ \omega(i, j) \geq \omega^{threshold} \\ 0, & f\ \ \omega(i, j) < \omega^{threshold} \end{cases} \quad \text{.....(21)}$$

Once the function $\varphi(i, j)$ is specified for the entire model grid, it can be used in the EnKF updating Eq. (5) in the similar manner as the additional weighting matrix $\boldsymbol{\omega}$. The modified equation for the EnKF updating that takes into account only part of the model grid adjacent to the well locations up to a certain distance is shown in Eq. (22).

$$\mathbf{M}(t_n) = \mathbf{M}(t_{n-1,n}) + \varphi \circ \boldsymbol{\omega} \circ \mathbf{K}(t_n) \cdot \left(\mathbf{D}(t_n) - \mathbf{H}(t_n) \cdot \mathbf{M}(t_{n-1,n})\right) \quad \text{.....(22)}$$

An example of the grid block retaining function converted from the weighting function at the threshold of $\omega^{threshold}$ = 0.3 for α = 1.0 is shown in Fig. 16. This threshold value can be correlated to the variogram range to ensure that only parts of the model that are significantly away from the wells

(approximately as the variogram range) are discarded from the updating step. The history matching results are compared in cases 10 and 11. There is no any improvement in RMSE of oil production when these two cases are compared to case 3 with a full covariance matrix consideration in the update step, but the computational time reduces and larger petroleum reservoir models can be handled with this approach. It is also advisable to use both grid block retaining function and matrix localization function for slightly better results (case 11 versus case 10).

Utica Shale Play

The shale reservoirs are of a significant importance for oil and gas industry in the United States and abroad (Lee and Kim 2016). The shale gas reservoirs contain significant reserves of natural gas primarily methane that is crucial to meet growing energy demand of the world in light of depleting conventional reserves and slowly growing share of renewable energy resources. The accurate construction of the reservoir model would allow developing the reservoir in an optimal manner. In this paper, we will look at Utica shale play of the Appalachian Basin, an asymmetric depression that extends from New York State in the north to Tennessee State in the south (US EIA 2017). The proven reserves are estimated at 6.4 trillion cubic feet of natural gas. We will apply EnKF history matching technique to the synthetized reservoir based on this Utica play to fine tune permeability model to honor historical production data in the form of the oil production rate. We assume same shape of the permeability distribution as come from the Utica log data. However, values of permeability are exaggerated by the factor of 10^6 to convert tight gas reservoir into conventional oil reservoir. Therefore, the accumulated hydrocarbons are assumed to be represented by light oil, and not gas.

Geological Description. Utica shale play consists of two Upper Ordovician organic-rich shale formations Utica and Point Pleasant, where the former formation overlies the latter one (US EIA 2017). They are part of Appalachian sedimentary basin system. The Koppe formation is deposited on the top of Utica formation. The Point Pleasant formation is deposited on the top of either Trenton or Lexington formations. The depths of Utica and Point Pleasant formations vary from 500 feet to 13,000 feet and from 500 feet to 14,000 feet below the surface, respectively, with a significant increasing trend from northwest to southeast of the play. The structure of the formations is controlled by the basement tectonics. These two formations span from New York State in the north to Tennessee State in the south. There are a number of major faults that run from northeast to southwest direction of the play and few minor faults that are perpendicular to major ones.

The geographic location of the area of interest in Utica and Point Pleasant formations is shown in Fig. 17. The more productive areas are characterized by 150 feet thickness of the combined Utica and

Point Pleasant formations. The Point Pleasant formation is considered more productive than Utica formation. Both formations contain an estimated 6.4 trillion cubic feet of natural gas.

From lithological point of view, Utica formation is primarily presented by calcareous shale with lower total organic carbon (TOC) in comparison with lower interval of Point Pleasant formation that is constituted of organic-rich shale with interbedded limestone and some siltstone. The shale content is higher in Utica formation that makes it is less susceptive to fracking procedures. These two factors, higher TOC and low shale content, determine higher productivity of lower interval of Point Pleasant formation. The thermal maturity of the reservoir rock increases from northwest to southeast of the play, which indicates higher chance of oil accumulation in the northwest part of the reservoir and higher chance of gas accumulation in the southeast part.

Objective of the History Matching. In our study, we will focus on the area of Utica shale play delineated by yellow polygon shown in Fig. 17, which is located in the eastern part of Ohio State. We have made a series of assumptions that are not typical for shale gas reservoirs. We will assume that only oil has been accumulated in the reservoir and permeability values are exaggerated by the factor of 10^6 to ensure reservoir fluid flow without fracking. There are five vertical production wells and two vertical injection wells that were drilled and completed in the area of interest. The wells are shown in Fig. 18. Red color is used to indicate production wells, and injection wells are colored in blue. The injection agent is water, which is used to maintain reservoir pressure and displace oil from the reservoir to the production wells. In this history matching study, we focused only on the Point Pleasant formation.

Because no production data are available, we generated a base case of permeability and porosity fields using log data from the available seven wells and then run a base case production scenario. The petrophysical models were simulated by using geostatistical technique SGS in DecisionSpace geomodeling software by Landmark, Halliburton. The permeability and porosity models are shown in Fig. 18 along with the well locations. The corresponding grid size is 100 x 100 x 6 blocks. Thus, there are total of 60,000 blocks, 59,664 of which are active. Each block size is 100.0 x 100.0 x 10.0 cube feet large. The standardized variogram model used for generating permeability values was previously presented in Eq. (9) with following substitution $h/a = \sqrt{\left(h_{x,y}/a_h\right)^2 + \left(h_z/a_v\right)^2}$. The variogram model is anisotropic with variogram range in horizontal direction being a_h = 3000.0 ft, while in vertical direction a_v = 10.0 ft. The histograms of permeability and permeability data and generated base case models are shown in Fig. 19. We can see that the distribution of the data is reproduced in the models.

The flow simulation was run on Nexus flow simulator by Landmark, Halliburton. The oil production rate and water/oil ratio are shown for each production well in Fig. 20 along with water injection rate for each injection well for fifteen years of oil extraction from the reservoir. The bottomhole pressure at production and injection wells is kept at 4000 psia and 8000 psia, respectively. The initial

water saturation was generated stochastically, where values vary between 0.18 and 0.78. The remaining pore volume is occupied by oil. No gas is present in the reservoir under in-situ conditions.

Therefore, the objective of this study is to incorporate all available relevant data to history match production data of the base case. We chose to calibrate only permeability model as it was done in the previously presented 11 cases. The potential data for the fine-tuning presented Utica-based petroleum reservoir model are oil production rate, water production rate (or oil/water ratio), and water injection rate. In order to perform history matching, EnKF has been chosen. The quality of the EnKF performance is judged based on the visual reproduction of the production data by the petroleum reservoir model and by calculating root mean square error (RMSE) values between permeability realizations and their truth. The RMSE values are computed as shown in Eq. (23). Here, *i*, *j*, and *l* are the indices of a permeability value *k* location in *X*, *Y*, and *Z* directions in the regular grid; N_X, N_Y, and N_Z are the number of the grid blocks in *X*, *Y* and *Z* directions; *r* is the realization index; k^r is the r^{th} permeability realization; k^{true} is the true permeability value; and *n* is the iteration index. The smaller RMSE values correspond to a better match between predicted permeability field and truth.

$$RMSE_{k^r} = \frac{1}{N_X N_Y N_Z}\sqrt{\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y}\sum_{l=1}^{N_Z}\left(k^r(i,j,l,n) - k^{true}(i,j,l)\right)^2}, \quad r = 1,\ldots,N_R \qquad (23)$$

In this case study, N_{m2} equals to $3.0 \cdot N_{PW} + 2.0 \cdot N_{IW}$ with $N_{PW} = 5$ and $N_{IW} = 2$ being number of production and injection wells, respectively; and 3 stands for oil production rate, water production rate and bottomhole pressure variables at production wells, while 2 stands for water injection rate and bottomhole pressure variables at injection wells.

Results of Application of the EnKF to Utica-based Reservoir. The synthesized Utica-based reservoir is set to produce for 15 years. The water injection is fixed at constant rate in both injection wells. By following guidelines from the previous section, at each yearly time step only the oil production rate data are assimilated into the model by means of the EnKF. After updating permeability model fifteen times according to the proposed methodology, the final permeability model is used to forecast oil production rate. The weighting function with power 1.0 and retaining function with threshold of 0.8 have been chosen as shown in Fig. 21 for covariance matrix localization to ensure the EnKF updating step is better-behaved. The same weighting function map was used for all vertical layers of the reservoir model under study. The ensemble consists of 100 realizations.

The results are presented in next several figures. The history matching of oil production rate for production well 1 and 4 is shown in Fig. 22. It is obvious that production data in black color are honored well and uncertainty of the prediction reduces as time progresses. The average of history matched results

in red are compared with the performance of the initial ensemble in grey if no production data were integrated into the model. The evolution of the first permeability realization is shown in Fig. 23 with RMSE values for all permeability realizations depicted in Fig. 24. It is obvious that permeability field improves as more production data are assimilated into the model. The initially overestimated permeability region highlighted by yellow circle is reduced in value through EnKF updating steps, while initially underestimated permeability region highlighted by red circle in increased in value. These changes make initial ensemble come closer to the permeability truth.

Therefore, the proposed data assimilation methodology with the EnKF has been proven as an efficient and robust procedure for history matching realistic petroleum reservoir based on Utica shale play geology.

Conclusions

In this manuscript, we presented a history matching of oil production rate for the synthesized petroleum reservoir based on Utica play geology using EnKF methodology for water flooding case, compared EnKF and EnKS on a small 2D example, and looked at a number of the modifications to EnKF and EnKS. The incorporation of oil production rate data at each yearly time step showed a significant improvement of the petroleum reservoir through the optimized history matching results of the production data and updated permeability ensemble that better resembles reality.

In order to implement EnKF or EnKS-based automatic history matching procedure, several improvements were made. The NST procedure was recommended and tested with successful results to avoid introducing bias to the highly skewed static model parameters that do not follow Gaussian distribution. The weighting function was applied to the EnKF and EnKS update step to avoid spurious covariance matrix and subsequent ensemble collapse. For the same reason, data error should not be ignored. The grid block retaining function was implemented to reduce computational time of the EnKF update step. Also, EnKS performed better in comparison to EnKF on a small 2D waterflooding example.

The examined methodology of EnKF or EnKS-based history matching and implemented improvements can be applied to any other type of the reservoir with confidence of getting great history matching results. The shale reservoirs that require fracking procedure are of the special interest for the proposed methodology testing, where the reservoir flow primarily happens via induced fracture network.

Future Work

The presented methodology should be extended to the lithological facies modeling. The spatial distribution of the facies is crucial, as they indicate better defined flow paths, barriers and baffle zones, and is integral part of the most geomodeling workflows. The facies distribution is of a categorical nature,

and should be treated slightly different from the distribution of continuous variables such as permeability or oil production rate.

The large volume time-lapse data such as 4D seismic, electromagnetic, and gravimetric data are widely collected during the repetitive surveys. Given large volume that is encompassed by these surveys, it is deemed feasible to incorporate these data type into a petroleum reservoir model for the improved permeability field estimation and, therefore, for the improved prediction of the reservoir performance behavior. The coarse resolution of these large volume data should be properly accounted for during the update step of the EnKF or EnKS.

Also, the speed of the flow simulation should be improved considerable to work with larger ensembles and bigger petroleum reservoir models. The distributed computing power (also known as cloud technology) for full flow simulation and replacement of the full physics of flow simulation with proxy models based on the machine learning techniques such as deep neural network (DNN) are suggested to reduce computational time at the forecast step of the EnKF and EnKS procedures. The parameter reduction techniques should be investigated to lower computational time of the update step of the data assimilation algorithms.

Acknowledgment

The financial support and permission to publish this manuscript provided by Landmark, Halliburton subdivision, are highly appreciated. We thank all colleagues, who were generous in reviewing this work, despite their busy time schedule.

References

Aanonsen, S.I., Nævdal, G., Oliver, D.S, Reynolds, A.C., and Vallès, B. 2009. The Ensemble Kalman Filter in Reservoir Engineering – a Review. *SPE Journal*, **14** (3): 393-412.

Chiles, J.-P. and Delfiner, P. 2012. *Geostatistics. Modeling Spatial Uncertainty: 2nd Edition*. New York: Wiley.

de Lima, A., Gentilhomme, T., Riffault, D., Allo, F., Anyzewski, A.S., Emerick, A.A. and dos Santos, M.S. 2017. Multi-Scale Ensemble-Based Data Assimilation for Reservoir Characterization and Production Forecast: Application to a Real Field. *Offshore Technology Conference Brasil, Rio de Janeiro, Brazil.*

Deutsch, C.V. and Journel, A.G. 1998. *GSLIB: Geostatistical Software Library and User's Guide: 2nd edition*. New York: Oxford University Press.

Evense, G., Hove, J., Meisingset, H.C., Reiso, E., Seim, K.S. and Espelid, O. 2007. Using the EnKF for Assisted History Matching of a North Sea Reservoir Model. *SPE Reservoir Simulation Symposium, Houston, Texas, USA.*

Evensen, G. 2009. *Data Assimilation – The Ensemble Kalman Filter: 2nd edition.* New York: Springer, 307.

Gilman, J.R. and Ozgen, C. 2013. *Reservoir Simulation: History Matching and Forecasting*. SPE Book Series, 120.

Goovaerts, P. 1997. *Geostatistics for Natural Resources Evaluation*. New York: Oxford University Press.

Haugen, V., Naevdal, G., Natvik, L.-J., Evensen, G., Berg, A.M. and Flornes, K.M. 2008. History Matching Using the Ensemble Kalman Filter on a North Sea Field Case. *SPE Journal*, **13** (4): 382–391.

Lee, K.S. and Kim, T.H. 2016. *Integrative Understanding of Shale Gas Reservoirs*. SpringerBriefs in Applied Sciences and Technology. Springer.

Lin, B., Crumpton, P. and Dogru, A. 2017. Field-Scale Assisted History Matching Using A Systematic, Massively Parallel Ensemble Kalman Smoother Procedure. *SPE Reservoir Simulation Conference, Montgomery, Texas, USA.*

Ma, X. Hetz, G., Wang, X., Bi, L., Stern, D., and Hoda, N. 2017. A Robust Iterative Ensemble Smoother Method for Efficient History Matching and Uncertainty Quantification. *SPE Reservoir Simulation Conference, Montgomery, Texas, USA.*

Mishra, S. and Datta-Gupta, A. 2017. *Applied Statistical Modeling and Data Analytics: A Practical Guide for the Petroleum Geosciences*. New York: Elsevier.

Patel, R.G.. Trivedi, J., Rahim, S. and Li, Z.. 20115 Initial Sampling of Ensemble for Steam-Assisted-Gravity-Draiange-Reservoir History Matching. *Journal of Canadian Petroleum Technology*, **54** (6): 424–441.

Pyrcz, M.J. and Deutsch, C.V. 2014. *Geostatistical Reservoir Modeling: 2nd edition*. New York: Oxford University Press.

Rwechungura, R., Dadashpour, M. and Kleppe, J. 2011. Advanced History Matching Techniques Reviewed. *SPE Middle East Oil and Gas Show and Conference, Manama, Bahrain*

US Energy Information Administration (EIA). 2017. *Utica Shale Play – Geology Review*. Independent Statistics & Analysis, US Department of Energy, 18.

Zhang, Y. and Oliver, D.S. 2011. History Matching Using the Ensemble Kalman Filter with Multiscale Parametrization: A Field Case Study. *SPE Journal*, **16** (2): 307–317.

Nomenclature

a	= variogram range, feet
a_h	= horizontal variogram range, feet
a_v	= vertical variogram range, feet
b	= distance function, feet
$\mathbf{C}$	= covariance matrix of the petroleum reservoir model
$\mathbf{C}_d$	= diagonal data covariance matrix
CV	= coefficient of variation
$\mathbf{D}$	= data matrix

d = data vector

F = cumulative probability distribution function of an arbitrary variable

f = flow simulation operator

G = cumulative probability distribution function of normally distributed variable

h = lag separation distance, feet

H = observation matrix

i = index of a location in *X* direction

j = index of a location in *Y* direction

K = Kalman gain matrix

k = permeability, mD

l = index of a location in *Z* direction

m = mean of variable

M = parameters of the petroleum reservoir model

M_1 = static parameters of the petroleum reservoir model

M_2 = dynamic parameters of the petroleum reservoir model

n = index of time step

N_d = number of the data

N_{IW} = number of injection wells

N_M = number of static and dynamic parameters of the petroleum reservoir model

N_{M1} = number of static parameters of the petroleum reservoir model

N_{M2} = number of dynamic parameters of the petroleum reservoir model

N_{PW} = number of production wells

N_R = number of realizations

N_X = number of the grid blocks in *X* direction

N_Y = number of the grid blocks in *Y* direction

N_Z = number of the grid blocks in *Z* direction

p = bottomhole pressure, Psi

q_{oil} = oil production rate, barrel/day

q_{water} = water production rate, barrel/day

r = realization index

RMSE = root mean square error

t = time, year

u = column vector of ones

w = index of the wells

x = coordinate in X direction

y = coordinate in Y direction

z = coordinate in Z direction

α = power in the additional weighting function

γ = variogram model

σ = standard deviation of variable

φ = grid block retaining function

ω = additional weighting function

List of Figures

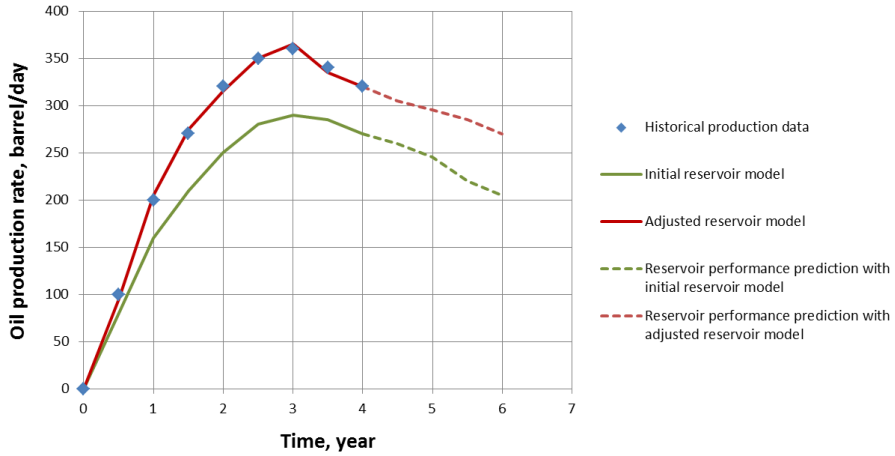


Fig. 1: Conceptual representation of history matching of a petroleum reservoir through oil production rate of a single well; blue dots represent recorded production data from the field, green line represents initial reservoir model based primarily on static parameter data (e.g. geologic data), red line represents adjusted reservoir model based on the updated static parameters that takes into account both static and dynamic parameter data; dashed sections of green and red lines are future predictions of the oil production – it is deemed that adjusted (history matched) model in red produces more accurate predictions of the oil production than unadjusted reservoir model in green, future production data can confirm this statement.

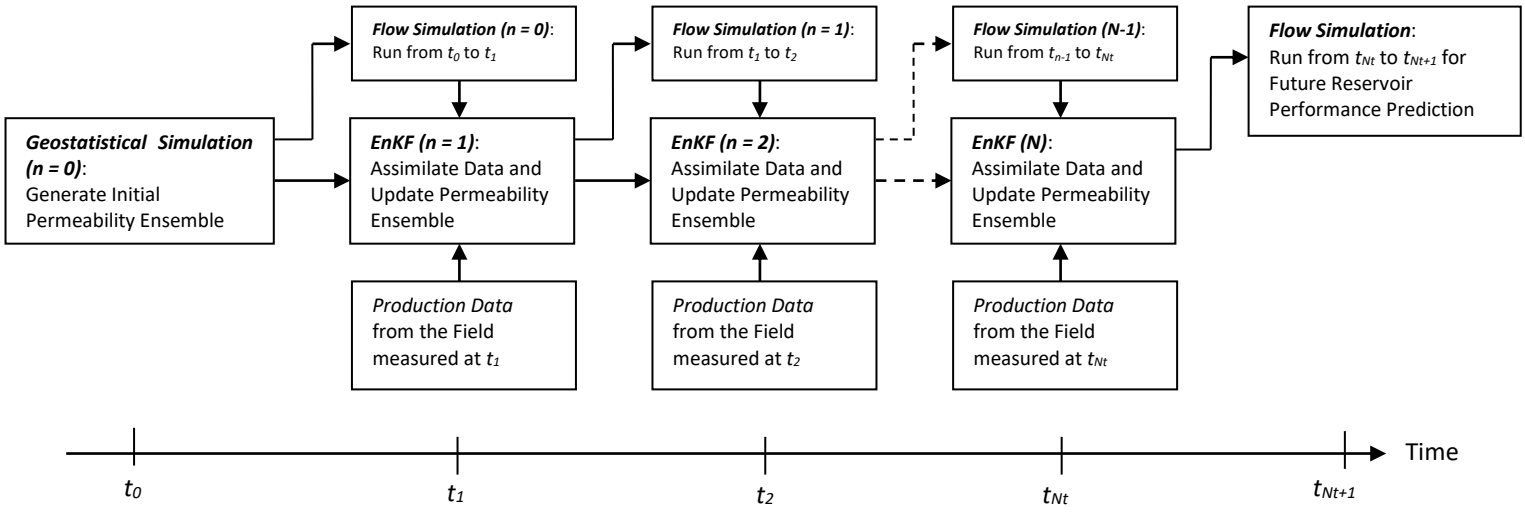


Fig. 2: Proposed workflow for history matching with EnKF for the Utica-based realistic petroleum reservoir

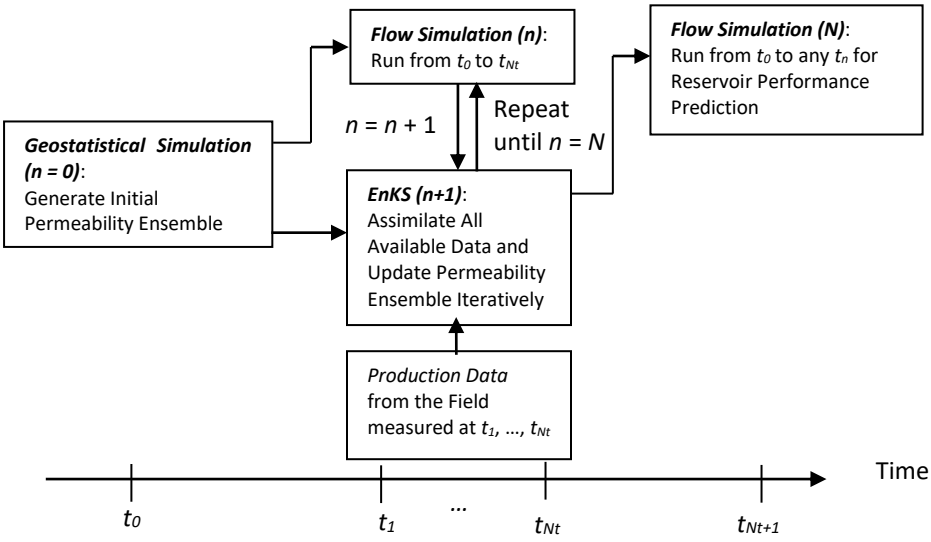


Fig. 3: Proposed workflow for history matching with EnKS for the Utica-based realistic petroleum reservoir

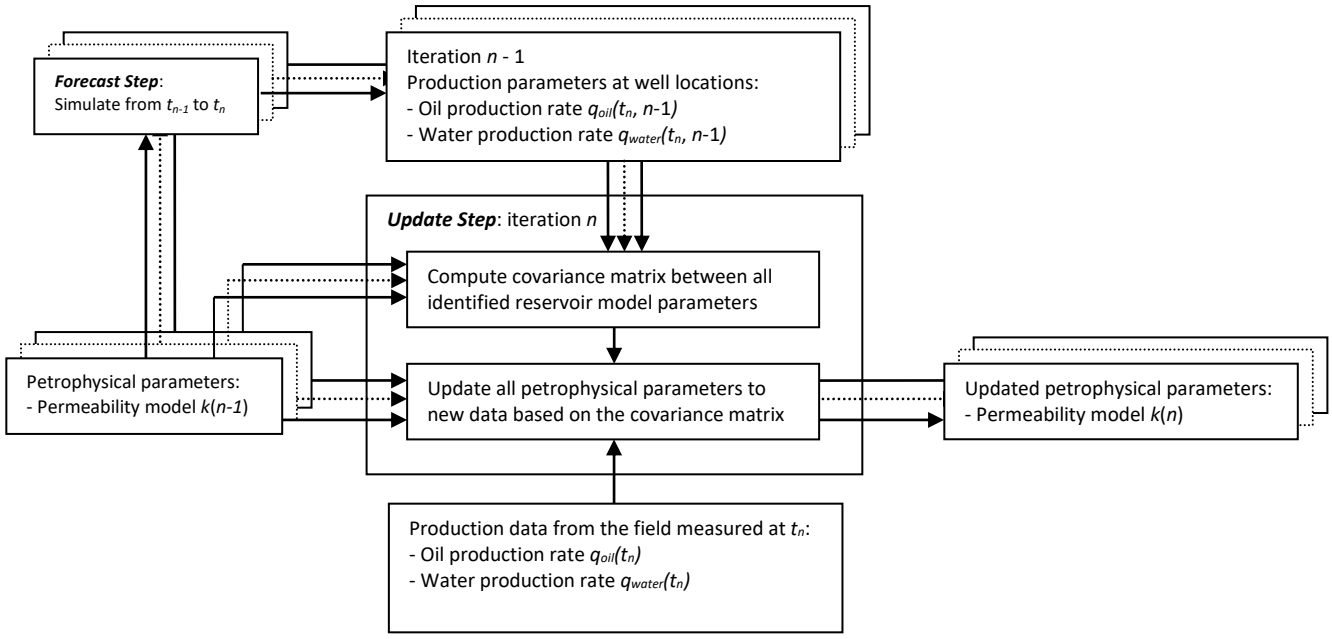


Fig. 4: Closer look at the update step of the EnKF. Note that an ensemble consists of multiple realizations N_R.

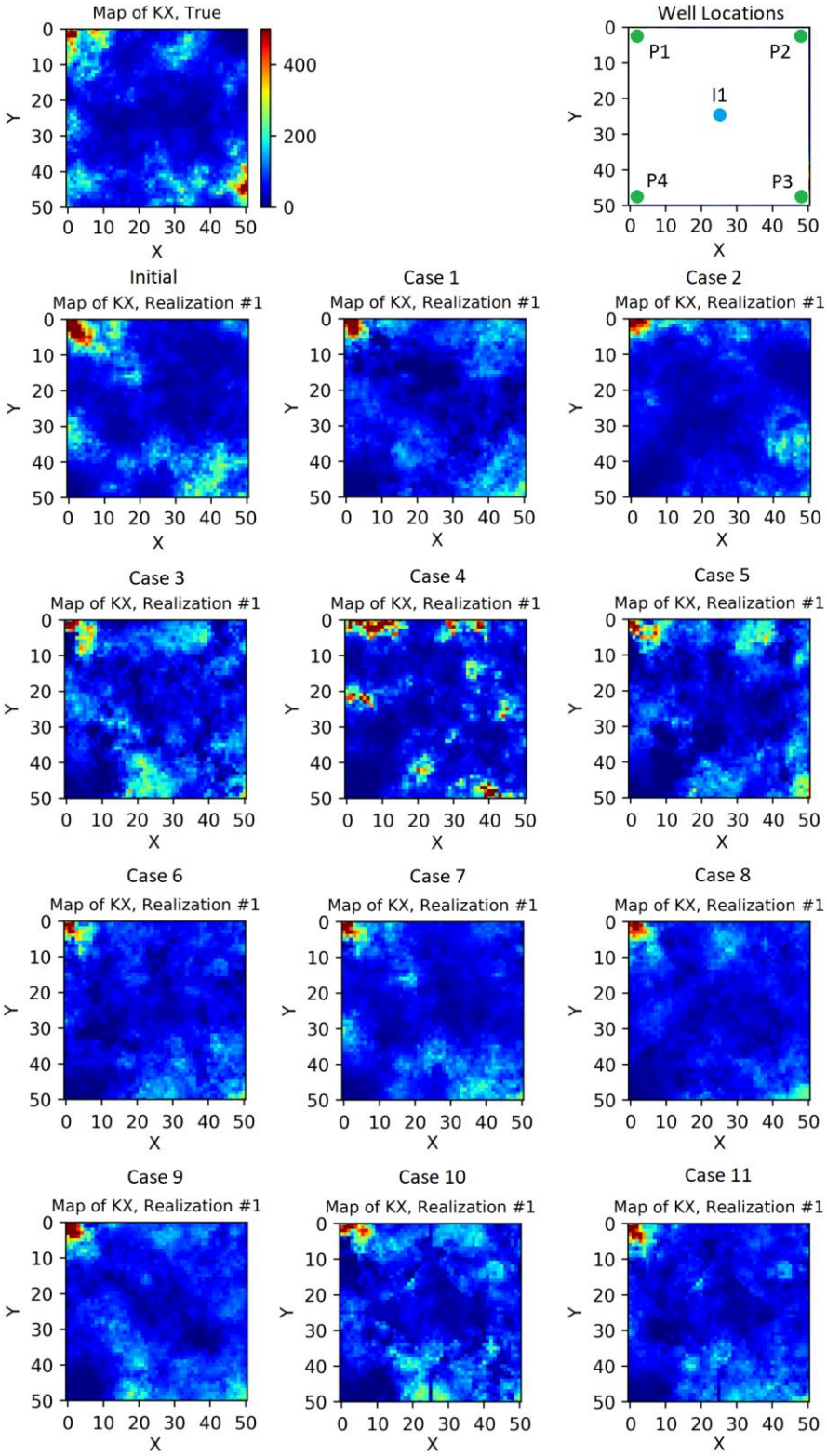


Fig. 5: 2D case study examples – map of true permeability field; well location map; map of first realization of the initial permeability field; and maps of first realization of permeability after final data assimilation iteration for each of 11 cases

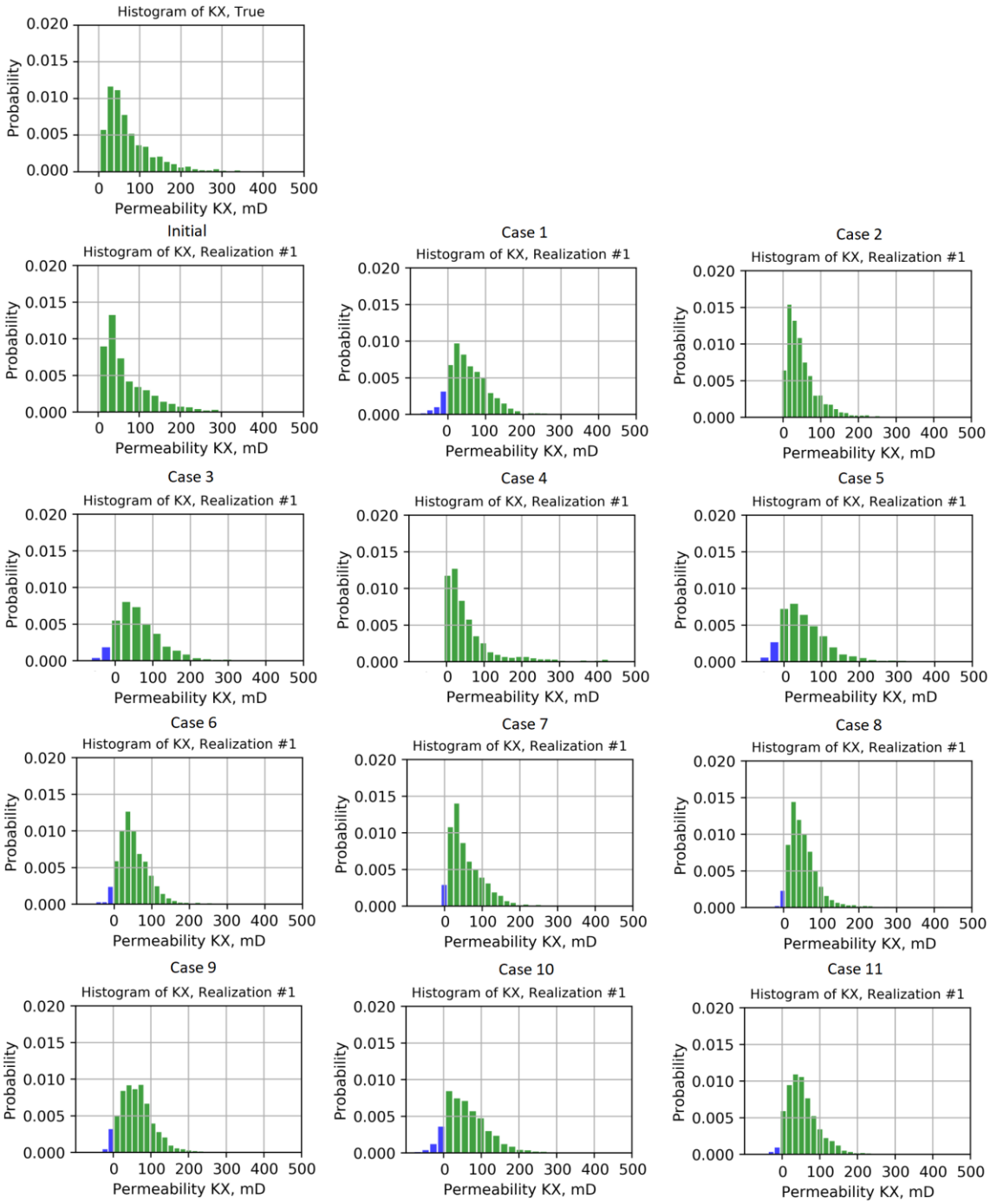


Fig. 6: 2D case study examples – histogram of true permeability field; histogram of first realization of the initial permeability field; and histograms of first realization of permeability after final data assimilation iteration for each of 11 cases

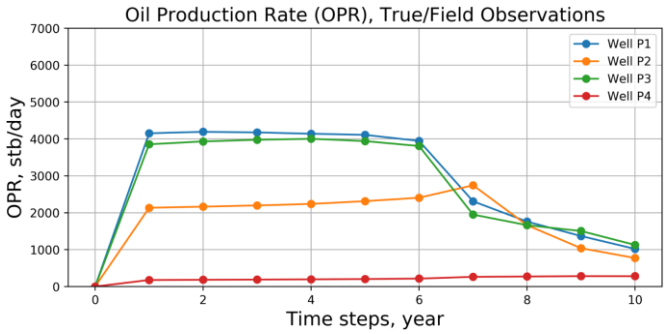


Fig. 7: 2D case study – recorded observations of oil production rate at four production wells

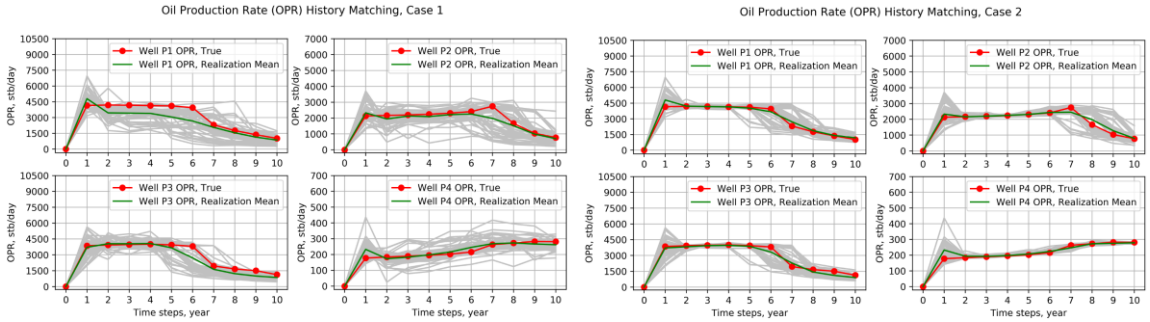


Fig. 8: 2D case study – history matching results for EnKF case 1 and case 2 for all 4 production wells

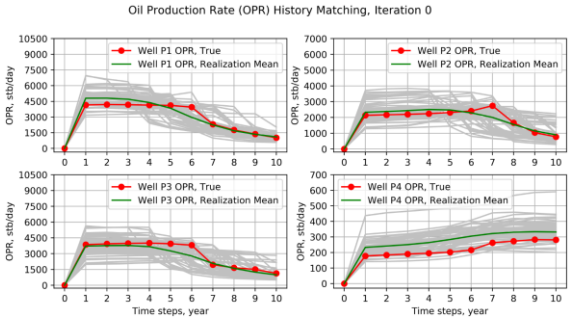


Fig. 9: 2D case study – initial oil production predictions for EnKS at four wells for cases 3 – 11

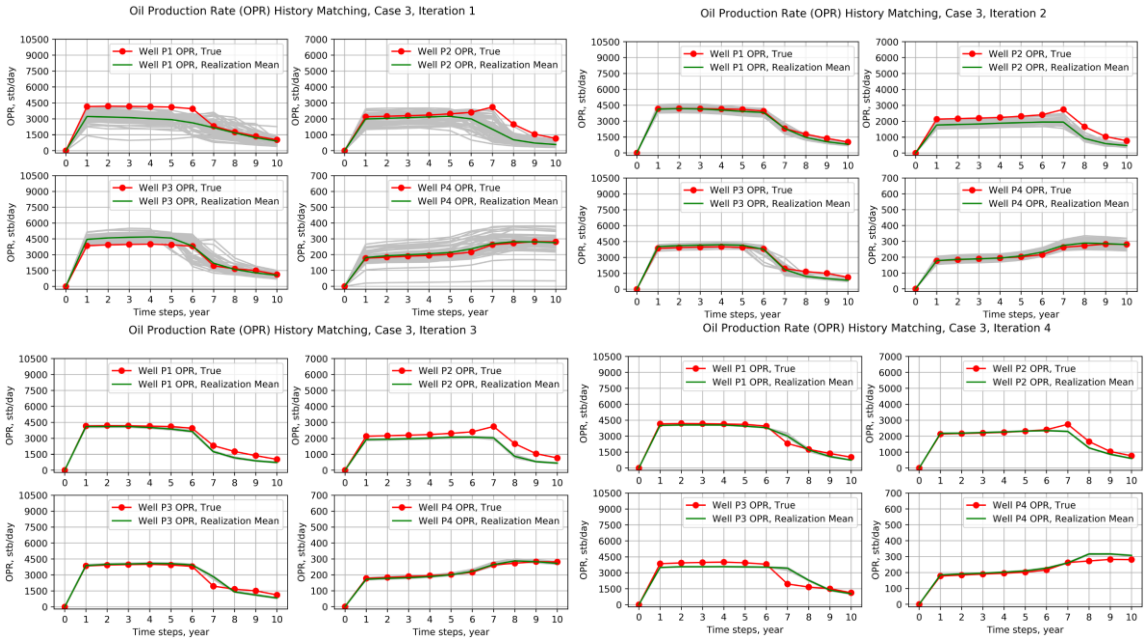


Fig. 10: 2D case study – history matched production predictions for EnKS iterations at four wells for case 3

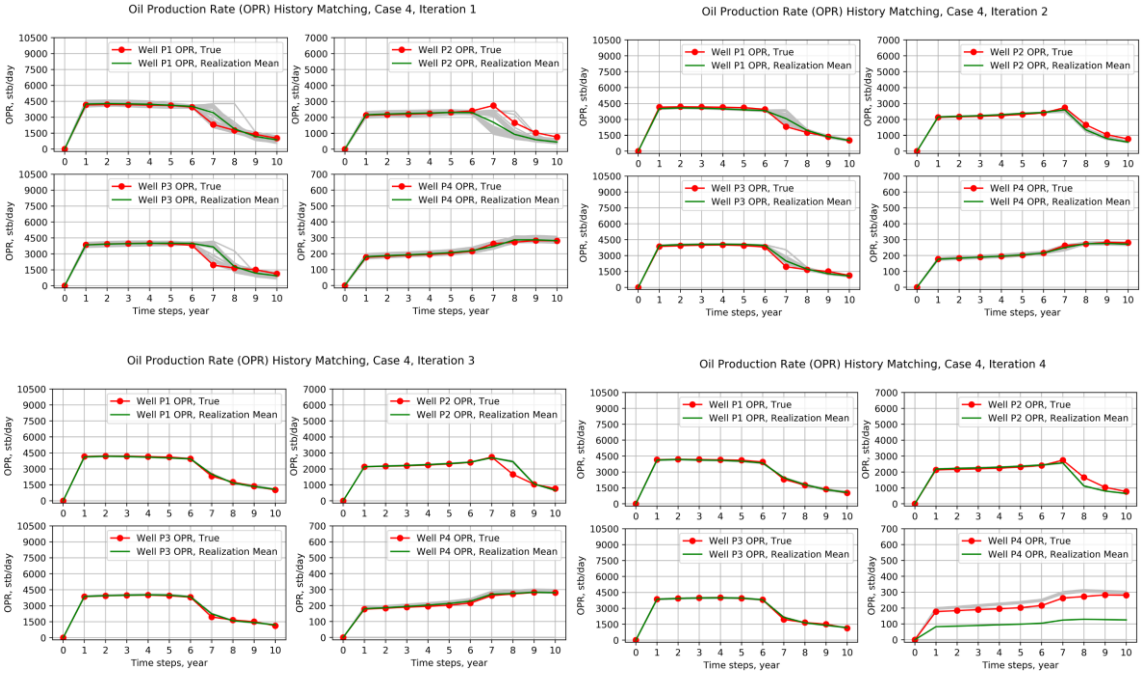


Fig. 11: 2D case study – history matched production predictions for EnKS iterations at four wells for case 4

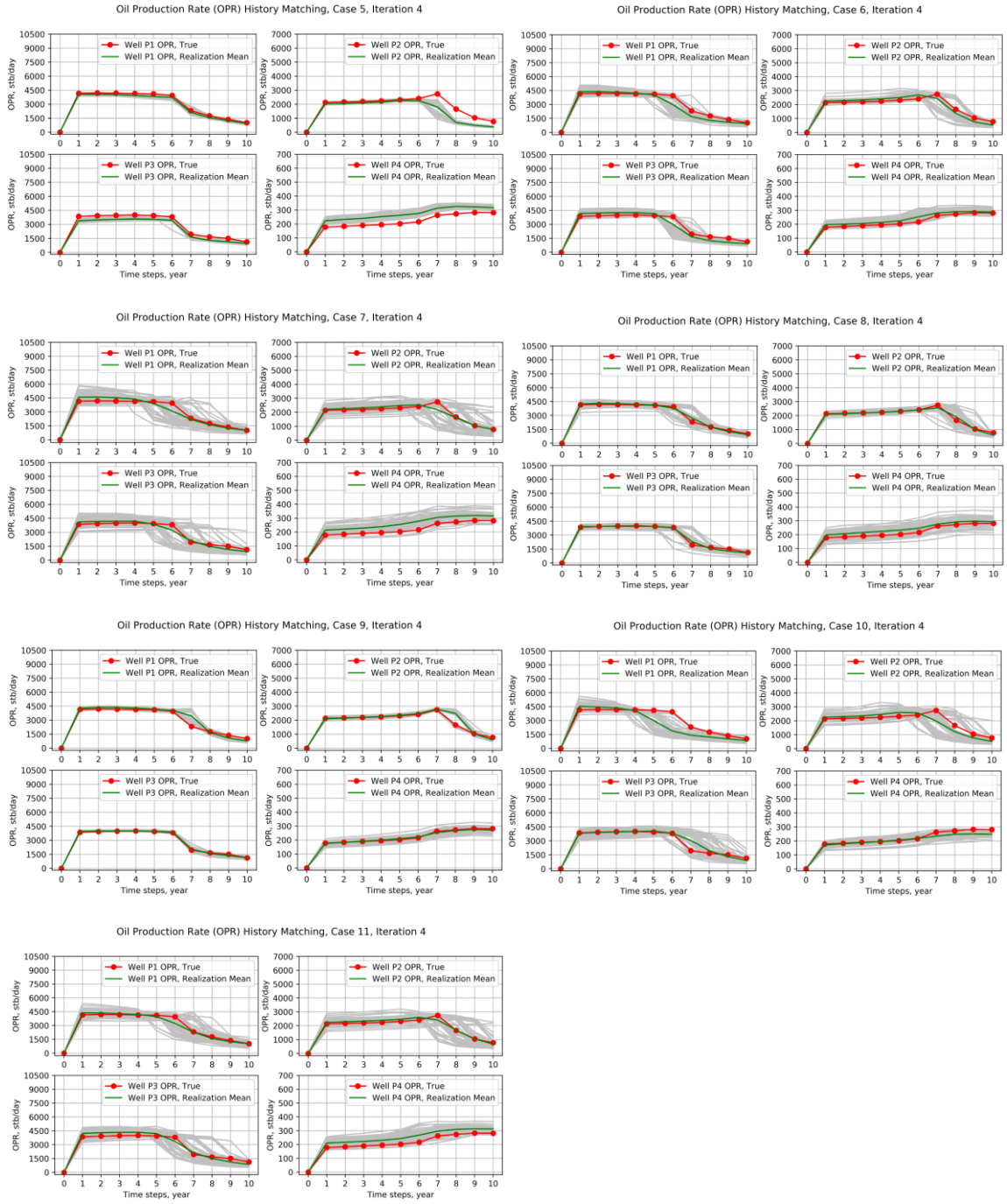


Fig. 12: 2D case study – history matched production predictions for final 4th EnKS iteration at four wells for cases 5 – 11

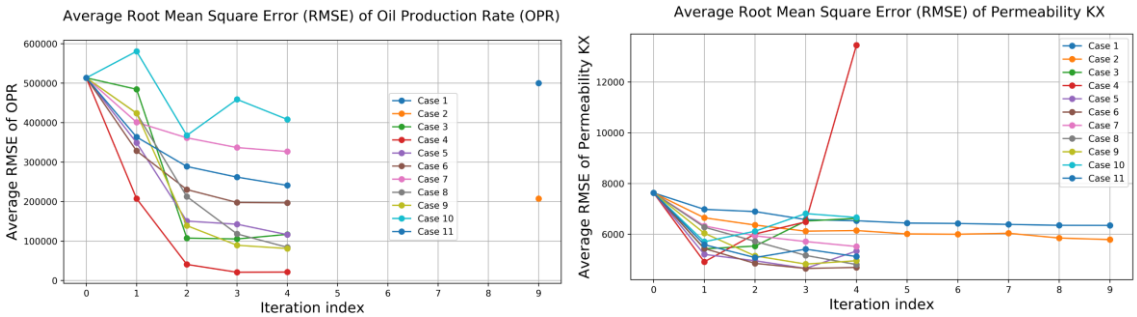
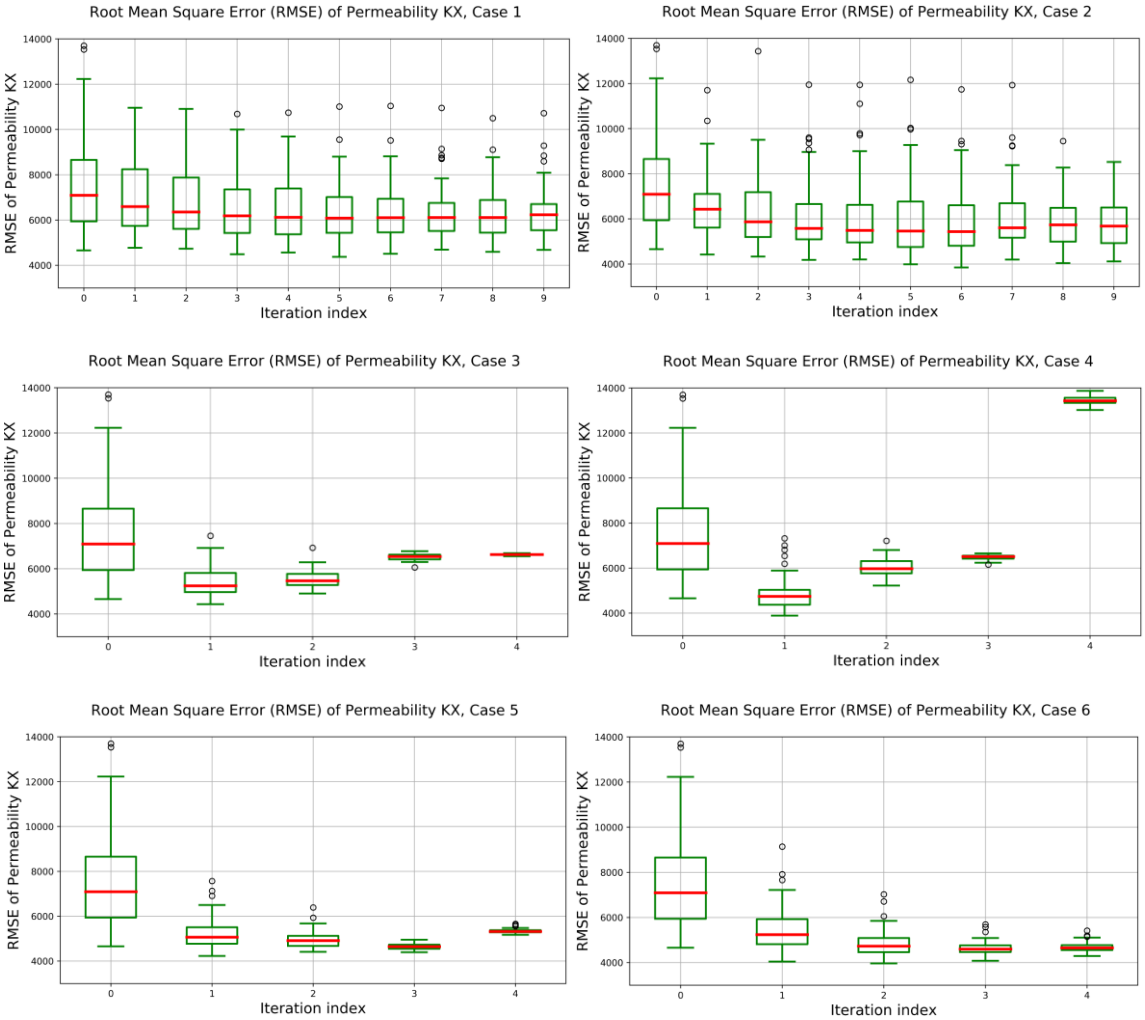


Fig. 13: 2D case study – average history matched RMSE values for EnKS iterations for all 11 cases



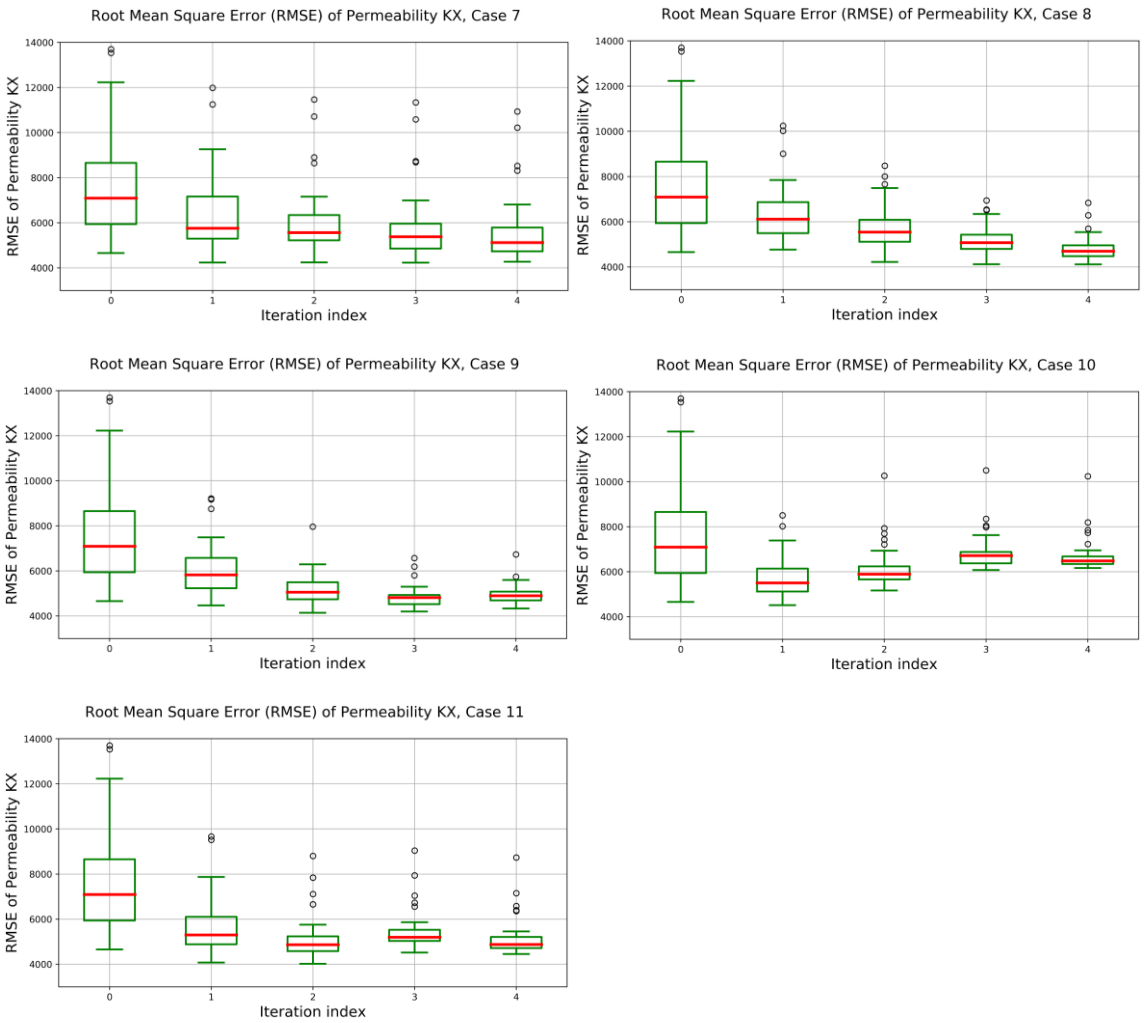
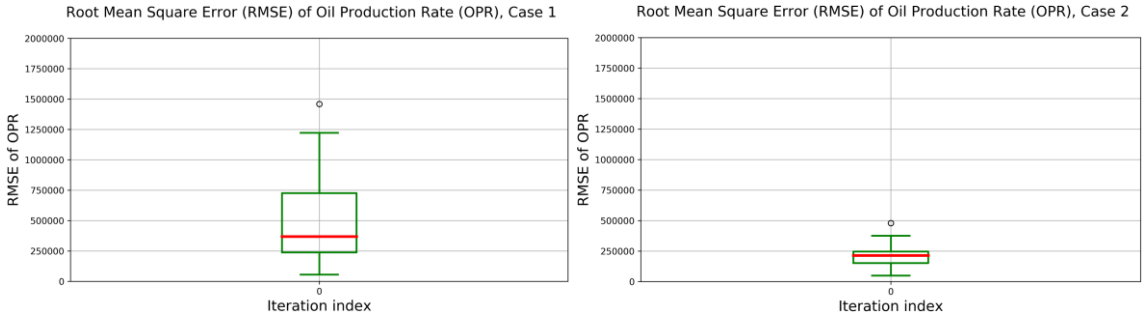


Fig. 14: 2D case study – RMSE values for permeability updates for all 11 cases



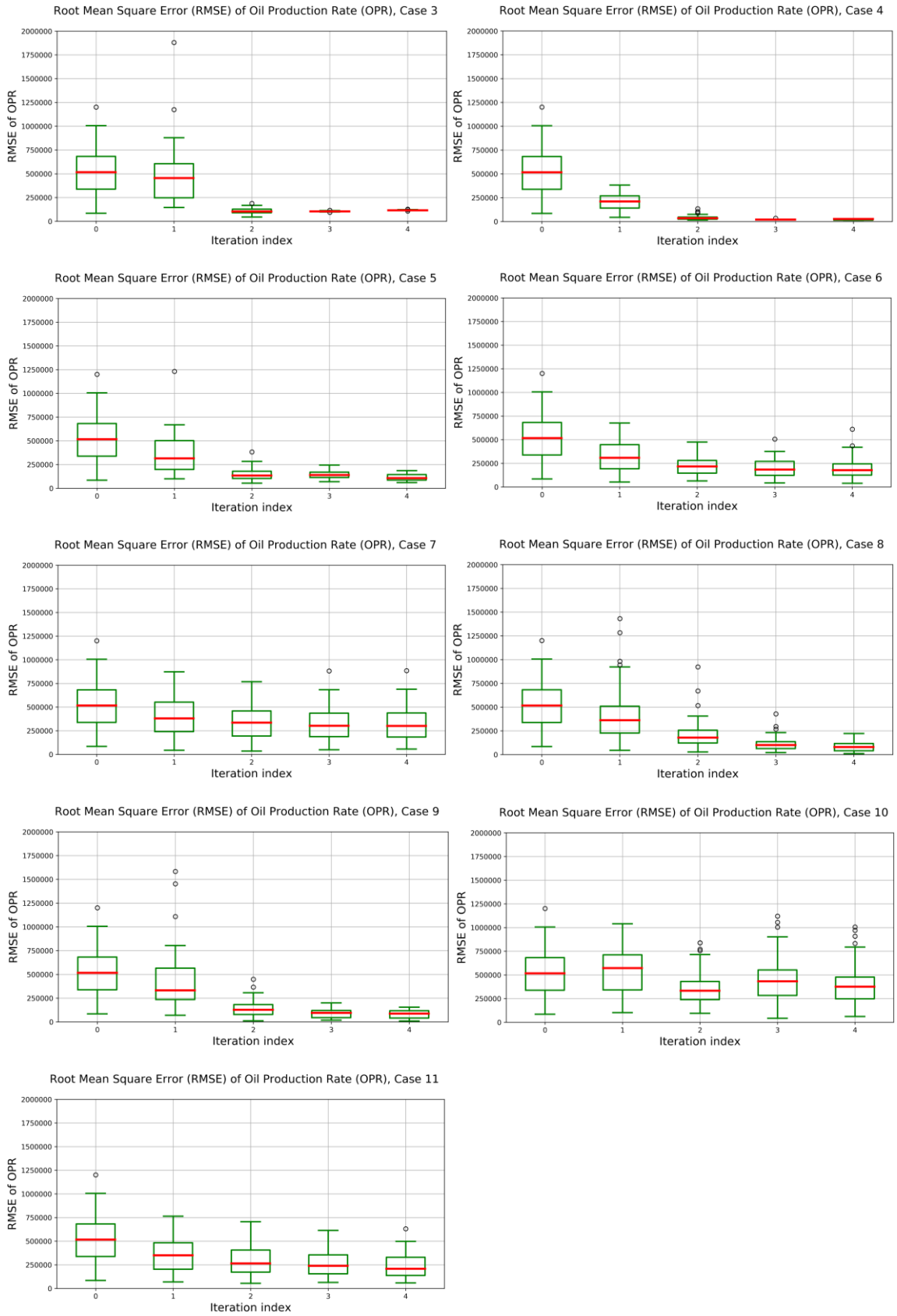


Fig. 15: 2D case study – RMSE values for oil production rate updates for all 4 wells and for all 11 cases

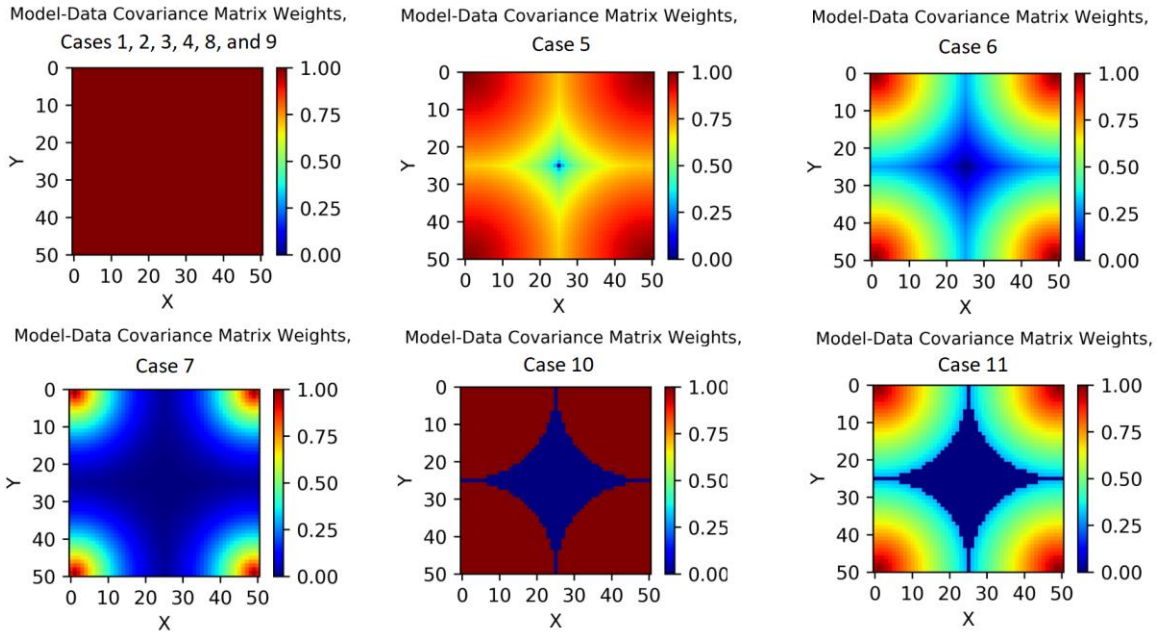


Fig. 16: 2D case study – maps of weighting functions either in the form of the covariance localization function, grid block retaining function or combination of these two functions for all 11 cases

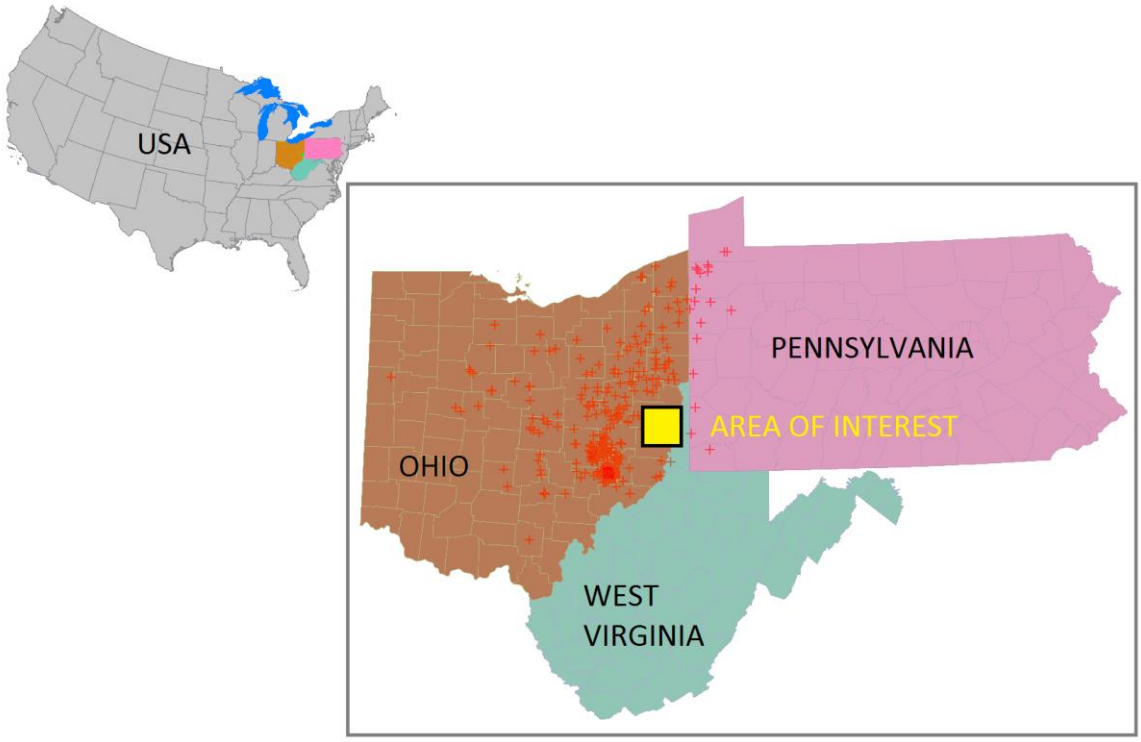


Fig. 17: 3D Utica case study – location of the area of interest on map of the United States with available well data shown as red crosses.

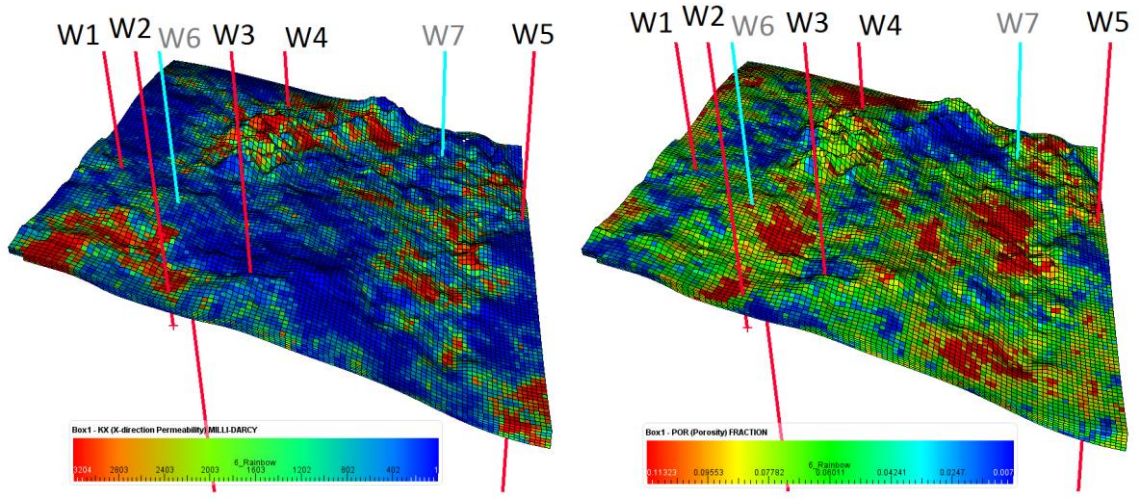


Fig. 18: 3D Utica case study – base case (treated as truth) permeability and porosity models of the Point Pleasant formation part of the Utica play reservoir under study. The vertical production and injection wells are shown as red and blue lines respectively.

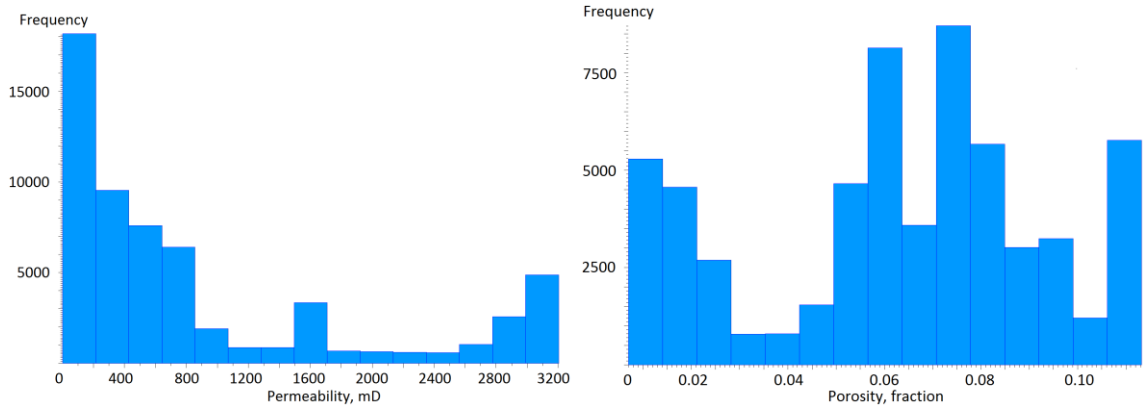


Fig. 19: 3D Utica case study – histograms of permeability and porosity models constructed from log data.

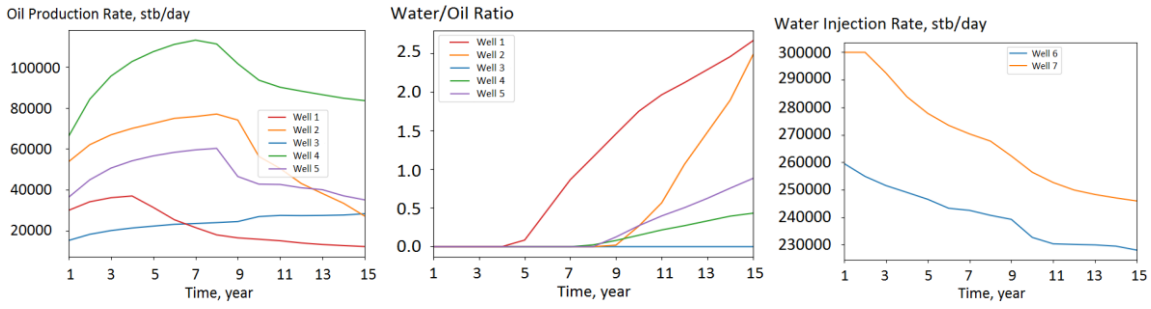


Fig. 20: 3D Utica case study – graphs of the recorded observations of oil production rate, water/oil ratio for each production well and water injection rate for each injection well for fifteen years of oil extraction from Utica play reservoir.

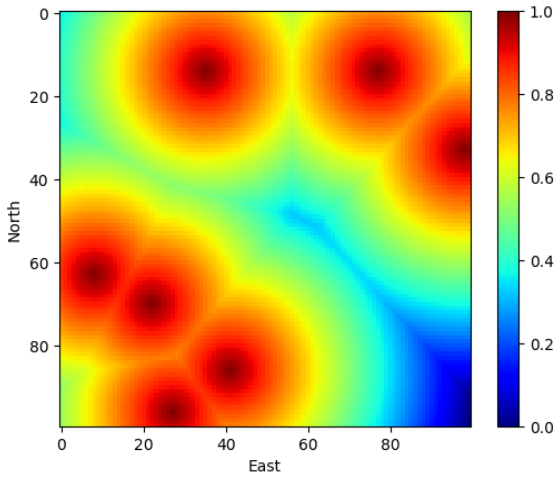


Fig. 21: 3D Utica case study – map of covariance localization weights based on the distance function as computed from Eq. (17) with power $\alpha = 1$.

Oil Production Rate at W1

Oil Production rate at W4

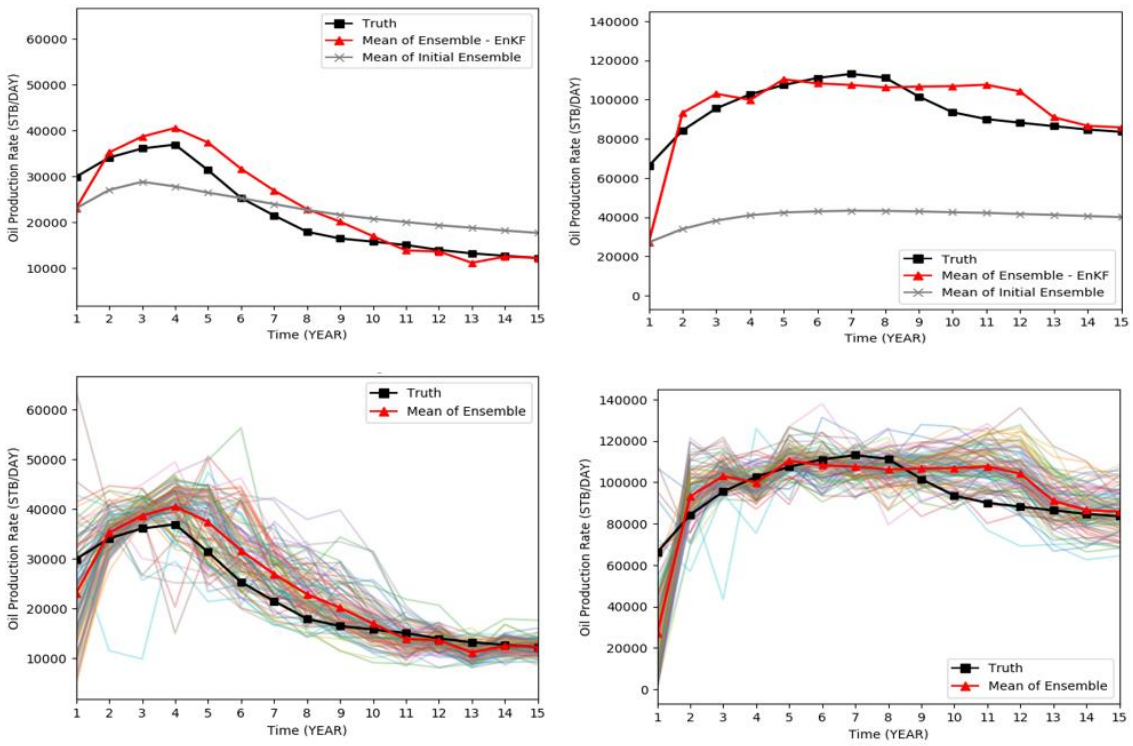


Fig. 22: 3D Utica case study – history matching results for wells W1 and W4 with entire ensemble members plotted at the bottom.

True permeability

Initial Permeability

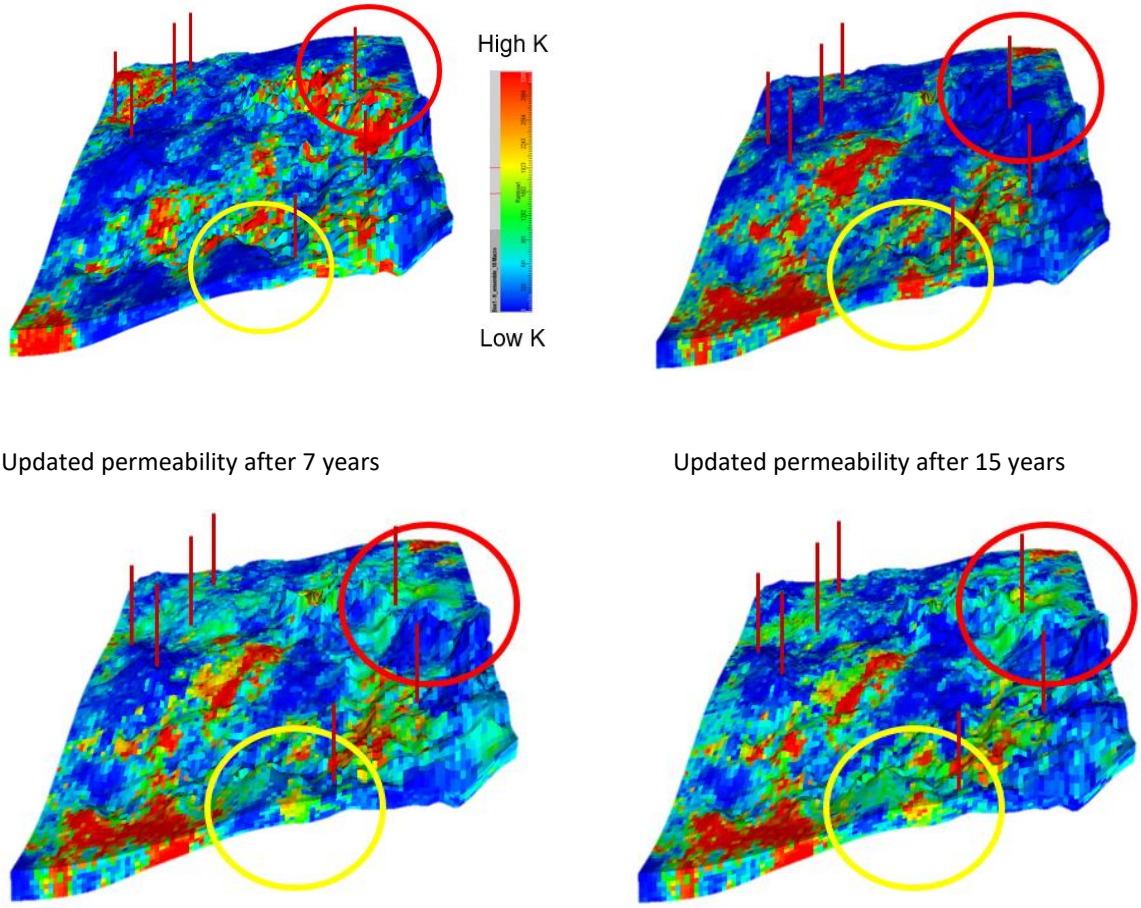


Fig. 23: 3D Utica case study – true, initial, and updated permeability fields. While overestimated permeability region in yellow circle undergoes permeability value reduction with EnKF, the underestimated permeability region in red circle undergoes permeability value increase with EnKF-based data assimilation history matching workflow.

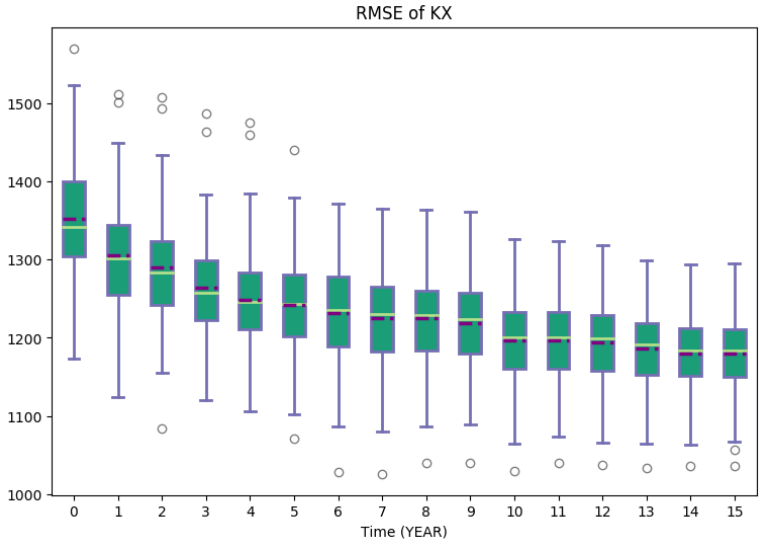


Fig. 24: 3D Utica case study – RMSE of updated permeability values. The permeability field becomes more similar to true permeability field as more production data are integrated into the reservoir with EnKF.

Biographies of Authors

Yevgeniy ZAGAYEVSKIY (yevgeniy.zagayevskiy@halliburton.com) is a product owner at Landmark, Halliburton, whose main responsibility involves supervision of product development for history matching, proxy flow modeling, field development optimization and uncertainty quantification. Yevgeniy holds PhD and MSc degrees in the Mining Engineering (Geostatistics) from the University of Alberta, and a BSc degree in the Petroleum Engineering from Kazakh-British Technical University, Kazakhstan, earned in 2015, 2012, and 2009 years, respectively. His PhD work is related to a multivariate grid-free geostatistical simulation. The master's thesis subject is about application of the EnKF in petroleum reservoir characterization. Yevgeniy has been an SPE member since 2010.

Hanzi Mao (hzmao@tamu.edu) is currently an intern at Landmark, division of Halliburton, and a PhD student in the Department of Computer Science and Engineering at Texas A&M University. She received her BEng and MEng in Telecommunication Engineering in 2009 and 2013 respectively, from Huazhong University of Science and Technology, China. Her research focuses on applying data science methods to explore the impact of heterogeneous biophysical factors on soil moisture dynamics under different hydro-climates.

Harsh Biren Vora (hv6@rice.edu) is currently an intern at Landmark, division of Halliburton, and a PhD student in Geophysics at the Department of Earth, Environmental and Planetary Sciences at Rice University. For his research, Harsh combines confined triaxial experiments measuring porosity, permeability and acoustic velocities, and Discrete Element modeling, to understand microfracture growth and fracture propagation mechanisms upon fluid injection. Harsh is the President of the AAPG Student Chapter at Rice for the 2017-2018 academic year, and aims to pursue a career as a geophysicist in the oil and gas industry. Harsh has received the BP-Rice Scholarship and Baker Hughes Scholarship during his time at Rice University.

Hui Dong (donghui@utexas.edu) is currently an intern at Landmark, division of Halliburton, and a PhD student in the Electrical and Computer Engineering Department at the University of Texas at Austin. He holds his BS degree from Tianjin University in 2010 and MSE degree from the University of Texas at Austin in 2012 both in Electrical Engineering. His current research includes nonlinear optical effect in nano-scale optoelectronic devices, optical force and acoustic force in opto-mechanical and acousto-mechanical systems, high-performance parallel computing, and computational electromagnetics. He has been a member of SPIE since 2013 and member of APS since 2015.

Terry Wong (terry.wong@halliburton.com) is a research scientist at Landmark, Halliburton, whose main task is development of petroleum reservoir numerical models. Terry holds an MSc degree in petroleum engineering from Stanford University and has been employed at Landmark since 1997. Terry has been a member of the SPE since 1979.

Dominic Camilleri (dominic.camilleri@halliburton.com) is a reservoir simulation technical advisor at Landmark, Halliburton, and has been involved in simulator development for 30 years. His current interests are simulation of multi-porosity and shale reservoirs using unstructured grids. He has an M.S. in Petroleum Engineering from the University of Texas at Austin in 1983 and a BSc in Mechanical engineering from the University of Malta in 1980.

Courtney Beck (courtney.beck@halliburton.com) is a Geologist/Earth Modeler with the Pinnacle Center of Excellence. Her current responsibilities include providing Earth Modeling for Pinnacle Integrated Sensor Diagnostic projects in the US. Courtney graduated with a B.S. in Geology from Texas A&M University in 2010 and earned a M.S. in Geological Sciences from the University of North Carolina at Chapel Hill in 2013. She joined Halliburton in 2013 and has worked on numerous exploration and asset development projects with oil and gas operators in the Utica shale, Eagle Ford shale, and East Texas stacked carbonate plays.

Charles Wang (charles.wang@halliburton.com) is a developer at Landmark, Halliburton, whose main responsibility involves development of Dynamic Frameworks to Fill. Charles graduated with a B.S. in Computer Science from University of Texas at Austin in 2005. He joined Halliburton in 2013 and has previously worked with the Earth Modeling team.