Characteristics of the EnKF for Geostatistical Problems

Yevgeniy V. Zagayevskiy, Amir H. Hosseini and Clayton V. Deutsch

Production data integration and history matching are important tasks in reservoir characterization and depletion planning. Automatic history matching in petroleum reservoirs is almost always associated with solving the inverse problem. A number of different inverse modeling techniques have been proposed in reservoir engineering and hydrogeology literature. In this paper, characteristics of the Ensemble Kalman Filter (EnKF) in solving the inverse problem for a single phase flow scenario in heterogeneous reservoirs are investigated. The EnKF is a simple and computationally cheap technique for estimation and prediction of model parameters and state variables. The paper focuses on EnKF implementation details in petroleum reservoir engineering and gives valuable recommendations for method employment. Pros and cons of the method are discussed to assess EnKF's capability and to define its application scope. The paper starts with theoretical background of EnKF and then presents a synthetic example with a number of sensitivity analyses. The method claims to be promising and powerful with a wide range of applications.

1. Background of the EnKF

The Ensemble Kalman Filter (EnKF) has evolved from the Kalman Filter (KF) named after Rudolf E. Kalman (1930 - present), a Hungary-American electrical engineer, college professor, and mathematical theorist. The KF is a recursive filter specially developed for dynamical models to remove noise content from the series of signal measurements and bring their estimated values closer to true ones (Kalman, 1960). Later the method has spread to other fields because of its universal properties, ease of application and satisfactory results. It is widely used in meteorology and oceanography for weather forecast (Daley, 1991) and ocean dynamics prediction (Bennet, 1992). Also applications can be found in hydrology (Houser et al., 1998), avionics and outer space vehicles as well. Since the KF is not able to manage large-scale problems, EnKF was devised as its numerical extension. KF techniques lack the ability to handle nonlinear models. For this reason Particle Filters (PF) can be used (Doucet, de Freitas and Gordon, 2001), which are claimed to be a generic form of EnKF.

The Kalman Filter is mostly applied to a dynamic model, which is characterized by model parameters (static variables) and state variables (dynamic variables describing state of a model) with some available data (Aanonsen et al., 2009). Similar to other inversion techniques, EnKF is prone to ill-posedness, that is, non-uniqueness of the results and instability of the predicted state variables with respect to changes in the observations. A dynamical model can be expressed in the following form (Equations (1) to (3)):

$$\frac{d\boldsymbol{X}(t_n)}{dt} = F(\boldsymbol{X};\boldsymbol{M}) + \boldsymbol{W}^{MP}(t_n)\text{, dynamical model} \tag{1}$$

$$\boldsymbol{X}(t_0) = \boldsymbol{X}_0\text{, initial state (boundary condition)} \tag{2}$$

$$\boldsymbol{D}_n = g_n\left(\boldsymbol{X}(t_n),\boldsymbol{M}\right) + \boldsymbol{\varepsilon}_n^D + \boldsymbol{v}_n \in \mathbb{R}^{N_D}\text{, observations} \tag{3}$$

where, $\boldsymbol{X}(t)$ is the vector of N_{SV} state variables; $\boldsymbol{M}$ is the vector of N_{MP} model parameters; F is the model operator that describes propagation of dynamical model in time t; n is the index for time t; $n = 0, \ldots, N_T$, $\boldsymbol{W}^{MP}$ is the model error vector caused by model structure only, which is also called white noise with zero-mean and unit variance; $\boldsymbol{D}_n$ is the data vector of available N_D measurements (observations) at time t_n; $\boldsymbol{\varepsilon}^D{}_n$ is the vector of measurement errors at time t_n; $\boldsymbol{v}_n$ is the perturbation vector added to the noisy data; and g_n is the observation operator that relates vectors $\boldsymbol{X}(t_n)$ and $\boldsymbol{M}$ with corresponding observations in time and space. Equation (1) describes the dynamical model, which propagates in time. Equation (2) defines boundary conditions for state variables, and Equation (3) is the observation operator.

The KF technique attempts to estimate true values of state variables and model parameters based on measurements and to predict $\boldsymbol{X}(t_n)$ at time steps t_n when no observations are available. If steady state model is considered, the method serves as an estimation tool only.

For the sake of simplicity, usually all parameters and state variables are combined together into one augmented vector $\boldsymbol{Y}_n$ (Equation (4)), from which all constituent vectors can be easily extracted through multiplication of $\boldsymbol{Y}_n$ by corresponding extraction matrix $\boldsymbol{H}^i$ (i = SV, MP, or D), which consists of only 0s and 1s. For instance, observation vector $\boldsymbol{D}_n$ is found in the manner shown in Equation (5).

$$\boldsymbol{Y}_n = \begin{bmatrix} \boldsymbol{X}_n \\ \boldsymbol{M} \\ g_n(\boldsymbol{X}_n, \boldsymbol{M}) \end{bmatrix} \in \mathbb{R}^{N_{SV}+N_{MP}+N_D} \tag{4}$$

$$\boldsymbol{D}_n = \boldsymbol{H}^D \cdot \boldsymbol{Y}_n + \boldsymbol{\varepsilon}_n^D \tag{5}$$

The solution to the problem is found in a two step procedure. First step is called a *forecast step* and is expressed in the form of model (prediction, forecast, or state) Equation (6). Here analyzed values of augmented vector $\boldsymbol{Y}^a_{n-1}$ at time t_{n-1} are forecasted to the next time interval t_n and new vector $\boldsymbol{Y}^f_n$ is obtained by using model operator F, which represents the dynamical model. In KF operator F is a linear (matrix form), but in EnKF the operator F could be a nonlinear as well. As it was stated before, $\boldsymbol{W}_{n-1}$ represents the model error.

The second step is an *analysis step* and is defined through observation (estimation, update, correction or analysis) Equation (7). Here forecasted values of augmented vector $\boldsymbol{Y}^f_n$ at step t_n are updated to $\boldsymbol{Y}^a_n$ by honouring available measurements $\boldsymbol{D}_n$. Also term Kalman gain $\boldsymbol{K}_n$ is introduced, whose expression can be found in Equation (8) . It is clear that covariance matrix $\boldsymbol{C}^f_n$ of forecasted variables $\boldsymbol{Y}^f_n$ should be known in order to conduct analysis step. Values of covariance matrix are derived from any prior information. Covariance matrix $\boldsymbol{R}_n$ of measurement errors is diagonal, since measurements of different variables are assumed to be independent. Moreover i^{th} measurement error ε_i should follow normal distribution with zero-mean and variance R_{ii} ($\boldsymbol{\varepsilon} \sim MVN(\boldsymbol{0}, \boldsymbol{R})$). Here observation matrix $\boldsymbol{H}_n$ is different from the extraction matrix $\boldsymbol{H}^i$, despite the fact that it also consists of only 0s and 1s. $\boldsymbol{H}_n$ relates variables from forecasted vector $\boldsymbol{Y}^f_n$ to their corresponding observations $\boldsymbol{D}_n$. Covariance matrix $\boldsymbol{C}^a_n$ of analyzed variables $\boldsymbol{Y}^a_n$ is updated through Equation (9).

Once analyzed values $\boldsymbol{Y}^a_n$ are obtained, they are used again for prediction of $\boldsymbol{Y}^f_{n+1}$ at next time step t_{n+1}. Recursive calculation process ends, when satisfactory results are found or no more new observations are available. Combination of forecast and analysis steps is usually called as an *assimilation step*, when newly available data are acquired, or just a *recursive step*, when no new data are available, but augmented vector $\boldsymbol{Y}^f$ is updated using old observations only.

$$\boldsymbol{Y}_n^f = F \cdot \boldsymbol{Y}_{n-1}^a + \boldsymbol{W}_{n-1}\text{, forecast step} \tag{6}$$

$$\boldsymbol{Y}_n^a = \boldsymbol{Y}_n^f + \boldsymbol{K}_n(\boldsymbol{D}_n - \boldsymbol{H}_n\boldsymbol{Y}_n^f)\text{, analysis step} \tag{7}$$

$$\boldsymbol{K}_n = \boldsymbol{C}_n^f\boldsymbol{H}_n^T(\boldsymbol{H}_n\boldsymbol{C}_n^f\boldsymbol{H}_n^T + \boldsymbol{R}_n)^{-1}\text{, Kalman gain} \tag{8}$$

$$\boldsymbol{C}_n^a = (\boldsymbol{I} - \boldsymbol{K}_n\boldsymbol{H}_n) \cdot \boldsymbol{C}_n^f \tag{9}$$

The main difference between EnKF and KF is that in EnKF the covariance matrices are calculated exclusively from the set of realizations that are simultaneously updated (Evensen, 1994). Thus, in EnKF the augmented vector $\boldsymbol{Y}_n$ is extended to matrix form shown in Equation (10):

$$\boldsymbol{Y_n} = \begin{bmatrix} x_{11,n} & x_{12,n} & \cdots & x_{1N_e,n} \\ \cdots & \cdots & \cdots & \cdots \\ x_{N_{SV}1,n} & x_{N_{SV}2,n} & \cdots & x_{N_{SV}N_e,n} \\ m_{11,n} & m_{12,n} & \cdots & m_{1N_e,n} \\ \cdots & \cdots & \cdots & \cdots \\ m_{N_{MP}1,n} & m_{N_{MP}1,n} & \cdots & m_{N_{MP}N_e,n} \\ g_n & g_n & \cdots & g_n \end{bmatrix} \in \mathbb{R}^{(N_{SV}+N_{MP}+1)\times N_e} \tag{10}$$

where, $x_{ij,n}$ is j^{th} possible value of i^{th} state variable at time t_n (i = 1, …, N_{SV}, j = 1, …, N_e, n = 1, …, N_T); $m_{kj,n}$ is j^{th} possible value of k^{th} model parameter at time t_n (k = 1, …, N_{MP},); N_{SV} is the number of state variables; N_{MP} is the number of model parameters; and N_e is the ensemble size or number of realizations.

Also KF Equations (6) to (9) are slightly modified for EnKF (Equations (11) to (14)). Now, model operator F resembles true nature of dynamical model and there is no more assumption of linearity in the model. Covariance matrix $\boldsymbol{C}^f_n$ is replaced by sample covariance matrix $\hat{\boldsymbol{C}}^f_n$ calculated from the ensemble of forecasted values $\boldsymbol{Y}^f_n$ (Equations (15) to (17)).

$$\boldsymbol{Y}_n^f = F(\boldsymbol{Y}_{n-1}^a) + \boldsymbol{W}_{n-1}\text{, forecast step}$$
(11)

$$\boldsymbol{Y}_n^a = \boldsymbol{Y}_n^f + \boldsymbol{K}_n(\boldsymbol{D}_n - \boldsymbol{H}_n\boldsymbol{Y}_n^f)\text{, analysis step} \tag{12}$$

$$\boldsymbol{K}_n = \hat{\boldsymbol{C}}_n^f\boldsymbol{H}_n^T(\boldsymbol{H}_n\hat{\boldsymbol{C}}_n^f\boldsymbol{H}_n^T + \boldsymbol{R}_n)^{-1}\text{, Kalman gain} \tag{13}$$

$$\boldsymbol{C}_n^a = (\boldsymbol{I} - \boldsymbol{K}_n\boldsymbol{H}_n)\cdot\hat{\boldsymbol{C}}_n^f \tag{14}$$

where

$$\hat{\boldsymbol{C}}_n^f = \frac{(\boldsymbol{X}_n^f - \overline{\boldsymbol{X}}_n^f)(\boldsymbol{X}_n^f - \overline{\boldsymbol{X}}_n^f)^T}{N_e - 1}\text{, sample covariance matrix} \tag{15}$$

$$\overline{\boldsymbol{X}}_n^f = \frac{(\boldsymbol{X}_n^f\cdot\boldsymbol{e}_{1\times N_e}^T)\cdot\boldsymbol{e}_{1\times N_e}}{N_e}\text{, mean of forecasted values} \tag{16}$$

$$\boldsymbol{e}_{1\times N_e} = \begin{bmatrix} 1 & 1 & \ldots & 1_{N_e} \end{bmatrix} \tag{17}$$

To summarize, the EnKF is a recursive method that integrates new observations into a model through specific assimilation procedure. EnKF is based on the set of realizations and consists of two steps: nonlinear forecast step, where model is propagated in time and relationship between variables is built, and linear analysis step, where all variables, despite whether their observations are available or not, are updated to honor available data through sample covariance matrix.

2. EnKF Algorithm

The EnKF algorithm consists of 3 main steps (Aanonsen et al, 2009):

(1) First, initialization step should be carried out. That is, initial or first-guess values of state variables $\boldsymbol{X}^a_0$ and model parameters $\boldsymbol{M}^a_0$ should be defined based on any prior knowledge. To generate a N_e realization set of

initial values of variables sequential Gaussian simulation with correct semivariogram model can be used. If semivariogram model is not chosen properly, EnKF results can lack for realistic geological pattern.

Usually, only initial values of primary variable are enough for EnKF initialization. Initial values of secondary variables are obtained at first forecast step (n = 1) through known model operator F.

(2) Once initial values of the variables are initialized, EnKF's recursive two-step mechanism can be launched. First, forecast step is performed (Equation (11)) and forecasted ensemble of augmented matrix $\boldsymbol{Y}^f_n$ for n = 1 is derived (prediction step). Usually model error $\boldsymbol{W}_{n-1}$ is assumed to be zero. More information on the model error can be found in (Evensen et al., 2007). Then, sample covariance matrix $\hat{\boldsymbol{C}}^f_n$ (n = 1) is calculated (Equations (15) to (17)), which is used in Kalman gain (Equation (13)). Finally, analysis step is run (Equation (12)) where $\boldsymbol{Y}^f_n$ is updated to $\boldsymbol{Y}^a_n$ (estimation step). That is, the difference between estimated values and corresponding observations are minimized. Other variables, whose measurements are not available, are updated to honor sample covariance matrix.

Recall that observation matrix $\boldsymbol{D}_n$ consists of ensemble of observations, where each realization is mean value $d_{i,n}$ plus some normally distributed error ε_{ij}.

(3) The procedure is repeated from step (2), where n takes value of 2. EnKF algorithm continues till satisfactory results are obtained or when newly acquired data are no more available ($n = N_T$).

Because of the presence of observation matrix $\boldsymbol{H}_n$ there is no need to calculate full sample covariance matrix. Instead only portion of it should be computed to get Kalman gain, what reduces computational requirements of the method. Therefore, the technique is proved to be relatively cheap and is able to handle large-scale systems.

3. EnKF Implementation Characteristics through Examples

Two step computational algorithm of the EnKF is simple, but in order to achieve results with high accuracy EnKF implementation details should be taken into account. For this reason EnKF characteristics is examined in the next paragraphs and some recommendations for implementation are presented. For this reason synthetic example with different cases is presented as below.

Description of the synthetic example

The main goal of the case study is to identify specific EnKF characteristics to favor better estimation results, i.e. to construct numerical model as close as possible to the reality. Despite EnKF is mostly devised for prediction of dynamical models, its properties are examined on a steady state synthetic example. The case study is focused on EnKF application on petroleum reservoir. However, it can be easily employed in other fields.

In this example two variable types are considered: bottom-hole pressure $\boldsymbol{P}$ (dynamical variable) and reservoir permeability $\boldsymbol{K}$ (static variable). The former variable type represents state variables of the reservoir model and the latter one – model parameters. Pressure and permeability values at each grid cells are considered as variables. The model is 100 units by 100 units with cell sizes of 2 units in each direction. The synthetic model is cell-centered and for reference case permeability and pressure are known everywhere over the grid (Figure 1).

Thus, 2500 (50 * 50 = 2500) variables of pressure stored in matrix $\boldsymbol{P}$ with N_e possible realizations and 2500 variables of permeability in matrix $\boldsymbol{K}$ that form reservoir model. Observations are extracted from permeability and pressure reference cases at observation locations (Figure 3) and are treated as observations $\boldsymbol{D}$ (data). Variables at other locations should be estimated. EnKF estimates are compared to the reference case to assess the precision of the method. Overall average difference (*OAD*) concept is used to quantify improvements in the estimation (Equation (18)). It is based on the overall average absolute difference between average estimate $\overline{Y}_{estimated;i,j}$ and true value $Y_{true;\,i,j}$ at each grid cell. $\overline{Y}_{estimated;i,j}$ is found by averaging all estimates lying in the domain of the moving average window of size 25 cells with a center in '*i, j*'th cell (Figure 2).

$$OAD = \frac{1}{N_{cells\ in\ X} \times N_{cells\ in\ Y}} \sum_{i=1}^{N_{cells\ in\ Y}} \sum_{j=1}^{N_{cells\ in\ X}} \left| \overline{Y}_{estimated;i,j} - Y_{true;i,j} \right|, \quad i = 1,\ldots,N_{cells\ in\ Y},\ j = 1,\ldots,N_{cells\ in\ X} \quad (18)$$

where $\overline{Y}_{estimated;i,j}$ is the average estimated value of a variable at *i*, *j* location over realizations; $Y_{true;\ i,\ j}$ is the true value of a variable at the same *i*, *j* location.

Lesser *AOD* indicates better match between estimates and true values. Also the results should be compared visually for pattern and spatial distribution reproduction. EnKF settings causing preferable outcomes are searched.

Different EnKF characteristics are examined on single isotropic synthetic example. Figure 6 shows permeability and pressure reference cases. Semivariogram map and semivariogram model are presented in Figure 5.

General findings

Implementation details of EnKF, such as proper ensemble size, number of assimilation steps, initialization values, measurement error rate, number of observations, are examined in this paper.

A) Influence of ensemble size (number of realizations) on EnKF estimation accuracy

To assess proper ensemble size N_e for acceptable EnKF estimation accuracy with computational efficiency, isotropic permeability field close to Gaussian distribution is chosen as the reference case (Figure 6). Pressure reference case is generated from permeability reference case by using flow simulator and boundary conditions. The boundary conditions are set to 3.5 units at North edge of the model and 2.0 units at South edge with no pressure gradients at East and West edges. Histogram plots of reference cases are shown in Figure 7. 25 observation wells spaced evenly on the field are used to obtain measurements (Figure 3). Permeability and pressure measurements are used in the inverse modeling. Observation wells #1, 4, 13, and 16 are used later to show convergence of estimates to true values with assimilation step number. Initial values of permeability are generated using sequential Gaussian simulation with correct semivariogram model (Figure 5). Its expression is shown in the form of Equation (19).

$$\gamma(h) = 0.05 + 0.04 \cdot Sph_{\substack{a_X=20.0 \\ a_Y=0.1}}(h) + 0.91 \cdot Sph_{\substack{a_X=20.0 \\ a_Y=23.0}}(h) \tag{19}$$

E-type or mean of EnKF estimation results with different realization numbers are shown in Figure 4. Despite that value of each cell is not reproduced exactly, regions of high and low values are captured well. Recall that number of observations is small relative to the total number of grid cells.

Graph of overall average difference (OAD) versus ensemble size for permeability and pressure estimation where permeability measurements are available and not are shown in Figure 6. It is clear that incorporation of additional data in the form of permeability observations changes EnKF results. Best match is obtained at 60 realizations for both cases (Figure 8).

From OAD comparison it is evident that estimates are getting better till 60 realizations and worsens a little bit for higher and lower realizations. When histograms of estimates and reference cases are compared, the best estimate is found at ensemble size of 60 realizations. Usage of smaller ensemble sizes causes chaotic results; large ones produce smooth estimates (compare maps and histograms on the Figure 6). Thus, it is recommended to use ensemble size around 60 for higher estimation accuracy. 60 realizations will be used as a default value throughout the paper.

Convergence characteristics of EnKF estimates at the four locations are shown in Figure 9 and Figure 10 for permeability and in Figure 11 and Figure 12 for pressure variables. Since pressure observations were used in the estimation process, all realizations with their mean values are converged to the true one in 2 assimilation step at most. However, exact reproduction is not achieved even for 10 recursion steps. We can see that despite wide spread of initial pressure values further estimated realizations are getting closer to the true values. For permeability variable convergence is quite different. Recall that inverse modeling does not seek for exact values; in contrast it only tries to evaluate model parameters that honor state variables' values. So, no unique solution exists.

B) Importance of number of assimilation steps for EnKF estimation results.

In the previous section it has been shown that about 1 assimilation step is enough for stable estimation results. Maps of updating vectors (portion after plus sign in the analysis Equation (12)) and corresponding variances at each assimilation step for the same permeability reference field with ensemble size of 60 realizations are shown in Figure 13 for permeability where permeability data are available and in Figure 14 for permeability where permeability data are not available. Apparently first one assimilation step is the most crucial one. Updating vectors at subsequent steps do not effect on EnKF estimates, since their mean and variance close to zero. From plots of overall average difference for both permeability and pressure (Figure 15) we can say that increase of assimilation step number positively impacts on permeability and pressure field estimation Also it is notable that uncertainty rate decreases with increase of assimilation steps. More assimilation steps produce better results. EnKF estimates are shown in Figure 16.

To sum it up, taking into concern convergence characteristics of estimates (Figure 9 to Figure 12), it is recommended to run at least 2 assimilation steps to obtain EnKF results with high accuracy. Interesting that estimation permeability results at 100th assimilation step when no permeability data are used is almost same as estimation permeability results at 1st assimilation step when permeability data are used.

C) Influence of measurement error on EnKF accuracy.

Same isotropic permeability reference field following Gaussian distribution is chosen to examine influence of pressure measurement error ε_{ii} on EnKF estimation accuracy. No permeability observations are used in estimation. Measurement error is expressed through its standard deviation, which is taken in percentage of corresponding observation. Graphs of OAD versus pressure measurement error in percents for permeability and pressure variables are shown in Figure 17. Measurement error varies from 0.001% to 50.00%. It is evident that OAD has the lowest value around 0.1% - 1.0% for both variables. However, since histograms are reproduced much better at error rate of 0.1%, it is chosen as a recommended one. Higher measurement error leads to smooth estimates. Thus, small measurement errors are desirable for better EnKF outcomes. Reference case and E-type of EnKF results for 0.001%, 0.1% and 50% measurement error rates are shown in Figure 18.

D) Effect of number of observations on EnKF estimation rate.

Number of pressure observations is varied between 1 and 49 in order to assess influence of number of observation on EnKF estimation results. Ne observation locations are shown in Figure 19. One observation at a time is added to EnKF to plot the curve OAD versus number of pressure observations used for the estimation. Numbers next to well locations represent the order in which observations are added as evenly as possible. The OAD versus number of observations is presented in Figure 20. It is clear that extra number of pressure observations improves estimation accuracy for both variables.

E) Choice of initial values.

Importance of proper initial permeability values is presented here. For this reason four different cases are examined using ensemble of size 60 realizations and pressure data: case 1 – random semivariogram model (Equation (20)) is used to generate first-guess values, no permeability data; case 2 – correct semivariogram model is used for initial values, no permeability data; case 3 – correct semivariogram model is used for initial values with same distribution range, no permeability data; case 4 – correct semivariogram model is used for initial values with same distribution range and permeability data.

$$\gamma(h) = 0.05 + 0.95 \cdot Sph_{\substack{a_X = 10.0 \\ a_Y = 10.0}}(h) \tag{20}$$

Estimation results of permeability are presented in Figure 21. Means of OAD are compared in Table 1. In all cases pattern of reference case is reproduced well with general rule that extra information for initialization of permeability field leads to better estimation accuracy, what is clear from average overall differences of different cases. However, it is a little bit different for permeability and pressure cases. Surprisingly any random semivariogram model may cause satisfactory results. But recall that, reference case is chosen close to normal distribution with zero mean and almost unit variance, what can make better results than they appear in reality. It is recommended to treat initial values with the highest possible caution.

Table 1 – Mean and standard deviation of Overall Average Difference (OAD) for four different cases

OAD	Case 1: Random Semivariogram Model	Case 2: Correct Model	Case 3: Correct Model + Correct Distribution Range	Case 4: Correct Model + Correct Distribution Range + Permeability Data
Permeability	1.110 and 0.030	1.080 and 0.360	1.110 and 0.050	1.060 and 0.028
Pressure	0.070 and 0.006	0.062 and 0.006	0.060 and 0.006	0.061 and 0.005

4. Pros and cons of the method

It has been shown that EnKF estimation and prediction technique for inverse modeling is very simple and relatively accurate; however, it has not only advantages, but disadvantages as well. The most important pros of the method are its simplicity, expressed in equations and algorithm, and easiness of implementation, as it does not need exclusive calculations of Jacobian matrix and it can be used as a black box (in comparison with SSC). There is no need to understand the mechanism of a model operator (flow process in our case). EnKF refers to it as to external independent process.

Also EnKF can be used both in inverse modeling, where observations of primary variable is not available. In examined cases we used pressure observation to estimate permeability and pressure fields from which good results are obtained. Also in some cases we added permeability data to conduct small sensitivity analysis.

One more good point of EnKF lies in its relatively low computational cost. Because of presence of the observation matrix $\boldsymbol{H}_n$ in analysis Equation (12), only part of sample covariance matrix is calculated indeed, what enormously saves time and enables to handle large-scale systems. Also ensemble size is relatively small, what facilitates computation process. Usually calculation time is constrained to the complexity of model operator (flow simulator). So, simplified model can be used instead, which describes dynamical process with same accuracy.

It should be mentioned that EnKF estimates are noisier then kriging ones, since it honors local values (no smoothing effect).

The biggest con of the method is its inability to handle highly nonlinear and/or non-Gaussian problems, since all estimates tend to be normally distributed and covariance matrix is not preserved in nonlinear systems. So the model can be simplified first and/or variable's value transformed to normal space or to logarithmical units in case of permeability estimation. However, since true distribution of an estimated variable is not known, the issue of back transformation still exists.

5. Conclusion

The Ensemble Kalman Filter has proven itself as a good estimation and prediction technique for inverse modeling with slightly nonlinear systems and Gaussian variables. It is computationally cheap and easy to implement. Some shortcuts can be applied to decrease computational time further while preserving same estimation and prediction accuracy. Some available approaches are covariance localization, square-root modification, and switching off model operator after first assimilation step.

As a general recommendation, the ensemble size should be larger than 60 realizations, at least 2 iteration steps and all prior knowledge in initialization of first-guess values should be used in order to obtain EnKF estimates with high accuracy.

Unfortunately, EnKF cannot handle highly nonlinear and non-Gaussian systems. If possible, the issues should be resolved in the future. Other promising work is in a search of way to decrease ensemble size (number of realizations), while preserving same quality by ranking realizations. It will help to save computational time, making EnKF less expensive. Also, simplified or proxy model can be designed to lower computational time of model operator (flow simulator in our case). Moreover, EnKF application to dynamical model should be examined to see its prediction capability. It would be interesting to apply EnKF to data coming from different scales. Finally, comparison of EnKF with other similar techniques, such as particle filter (PF), which is more general estimation and prediction method, or unscented Kalman filter (UKF), should be carried out.

References

Aanonsen, S.I., Nævdal, G., Oliver, D.S., Reynolds, A.C. and Valles, B., 2009, The Ensemble Kalman Filter in Reservoir Engineering - a Review, *SPE Journal*, 393-412.

Bennett, A.F., 1992, *Inverse methods in physical oceanography,* New York: Cambridge University Press.

Daley, R., 1991, *Atmospheric data analysis,* New York: Cambridge University Press.

Doucet, A., de Freitas, N. and Gordon, N., 2001, *Sequential Monte Carlo Methods in Practice,* New York: Statistics for Engineering and Information Science: Springer.

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

Evensen, G., 1994, Sequential Data Assimilation with a Nonlinear Quasi-geostrophic Model Using Monte Carlo Methods for Forecast Error Statistics, *Journal of Geophysical Research*, 99, 10,143-10,162.

Houser, P.R, Shuttleworth, W.J., Famiglietti, J.S., Gupta, H.V., Syed, K.H. and Goodrich, D.C., 1998, Integration of soil moisture remote sensing and hydrologic modeling using data assimilation, *Water Resources Research*, 34, 3405-3420.

Kalman, R.E., 1960, A new approach to linear filtering and prediction problems, *Journal of Basic Engineering (Citeseer)*, 82, 35-45.

Figures

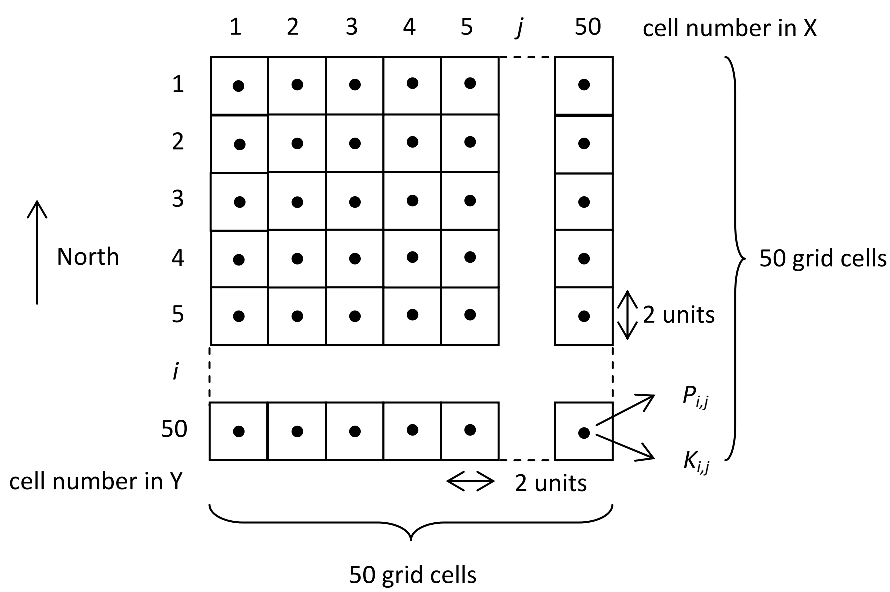

Figure 1 – Synthetic 2D case study. Numerical reservoir model with a grid.

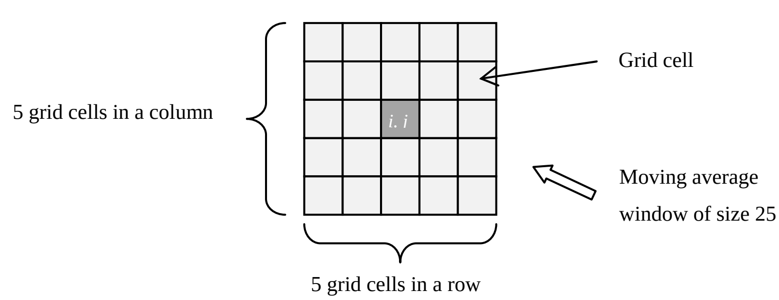

Figure 2 – Definition of size of moving average window.

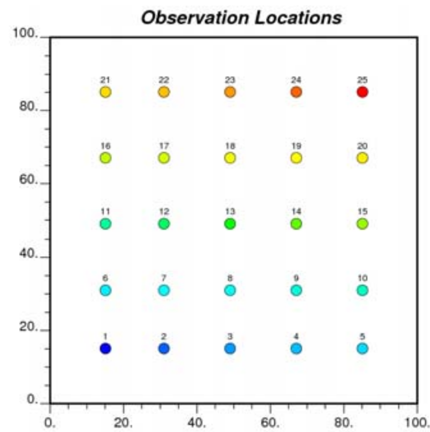

Figure 3 – Observation locations and their order.

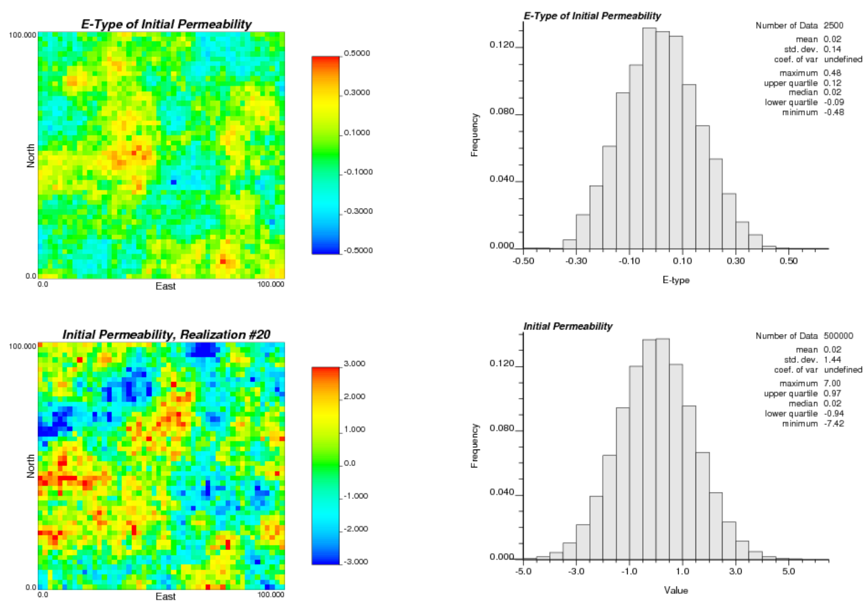

Figure 4 – E-type map and single realization # 20 of initial permeability with corresponding histograms.

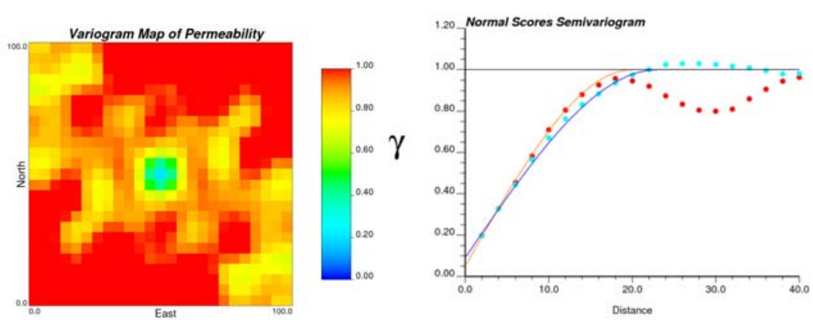

Figure 5 – Semivariogram map and semivariogram model in *X* and *Y* directions of permeability reference case.

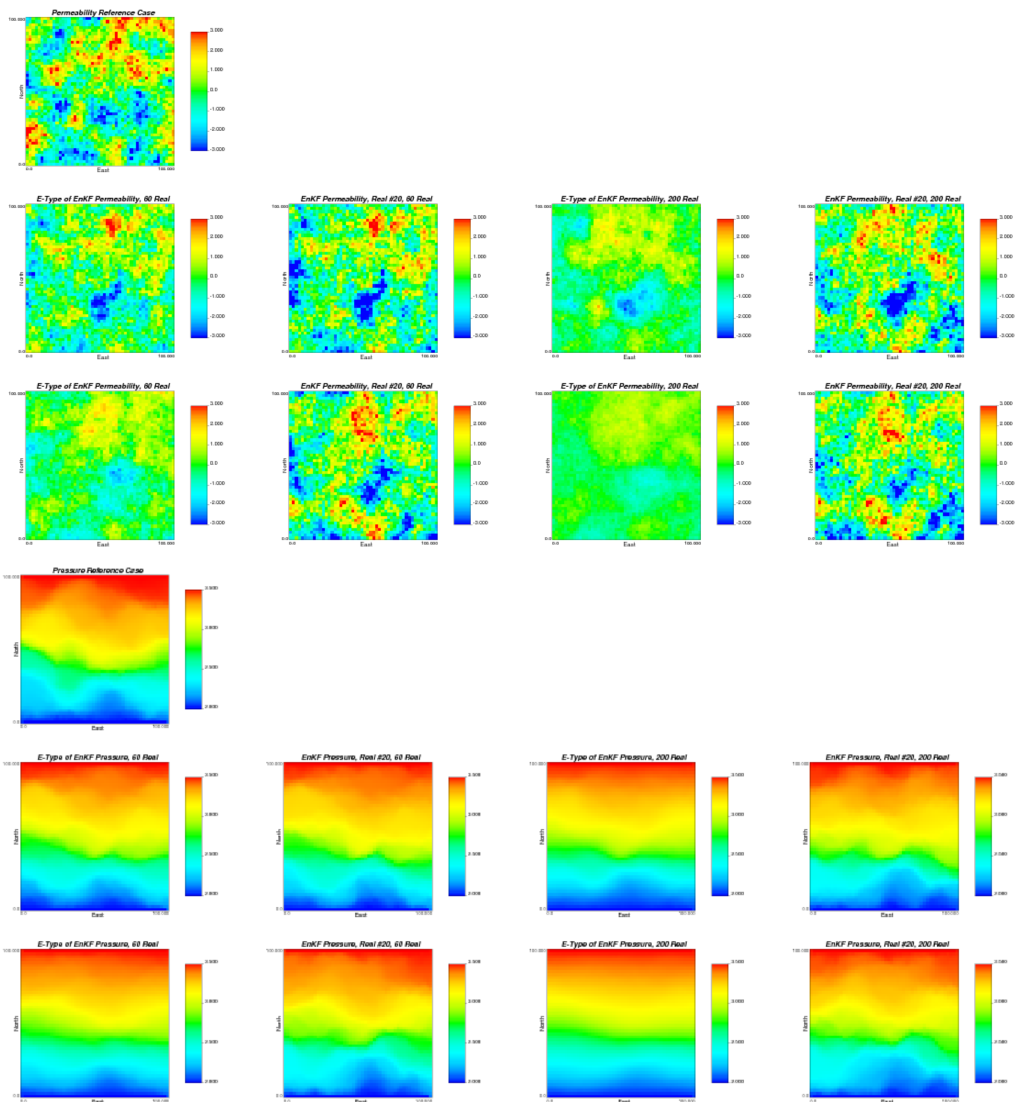

Figure 6 – Permeability and pressure reference cases and EnKF permeability and pressure estimates. Row 1 – permeability reference case; row 2 (from left to right) – E-type of EnKF permeability estimates using 60 realizations and permeability data, single realization of EnKF permeability estimates using 60 realizations and permeability data, E-type of EnKF permeability estimates using 200 realizations and permeability data, single realization of EnKF permeability estimates using 200 realizations and permeability data; row 3 is same as row 2, but no permeability data are used; row 4, 5, and 6 are same as row 1 – 3, but pressure results.

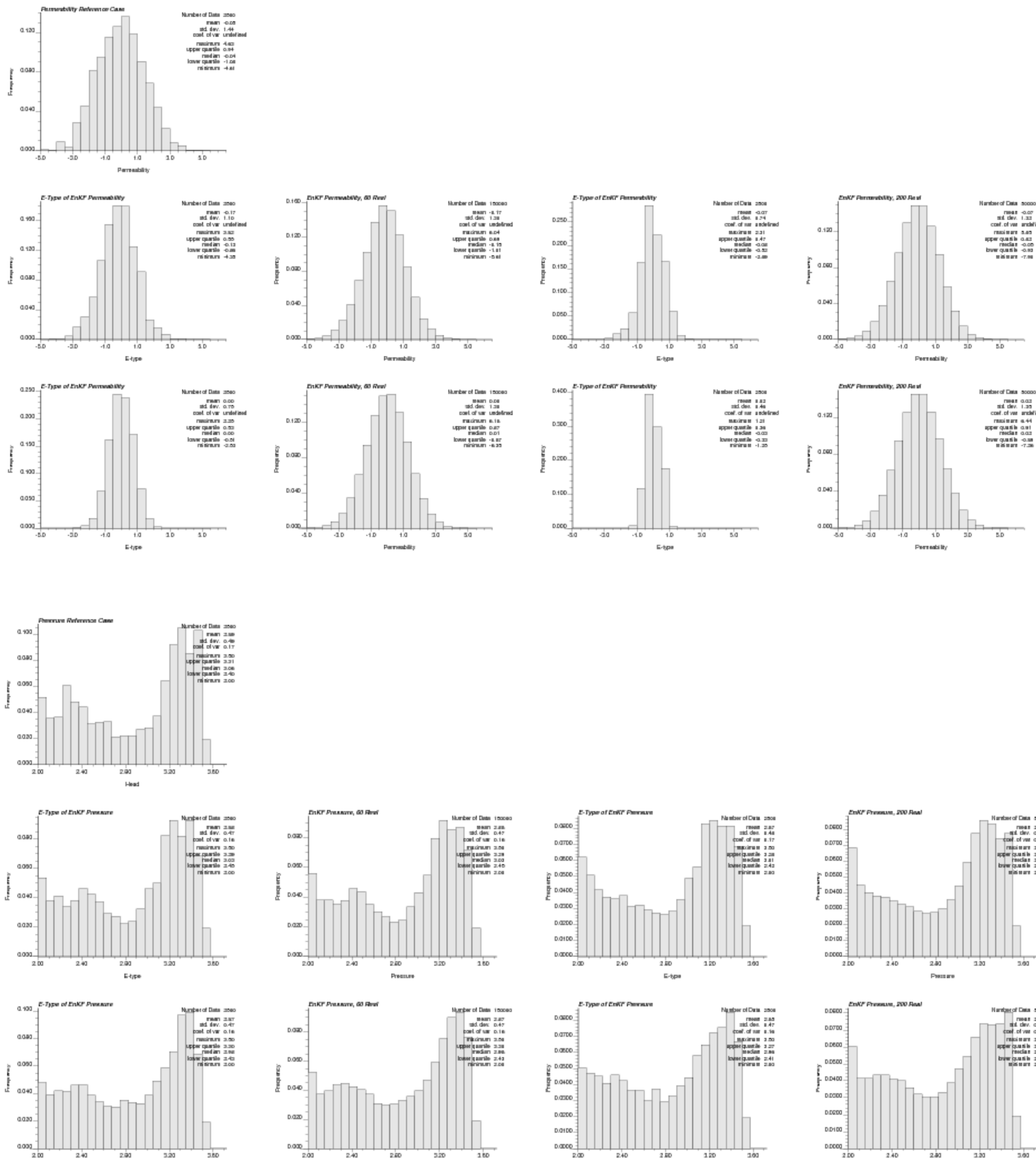

Figure 7 – Histograms of permeability and pressure reference cases and histograms of EnKF permeability and pressure estimates. Row 1 – permeability reference case; row 2 (from left to right) – E-type of EnKF permeability estimates using 60 realizations and permeability data, single realization of EnKF permeability estimates using 60 realizations and permeability data, E-type of EnKF permeability estimates using 200 realizations and permeability data, single realization of EnKF permeability estimates using 200 realizations and permeability data; row 3 is same as row 2, but no permeability data are used; row 4, 5, and 6 are same as row 1 – 3, but pressure results.

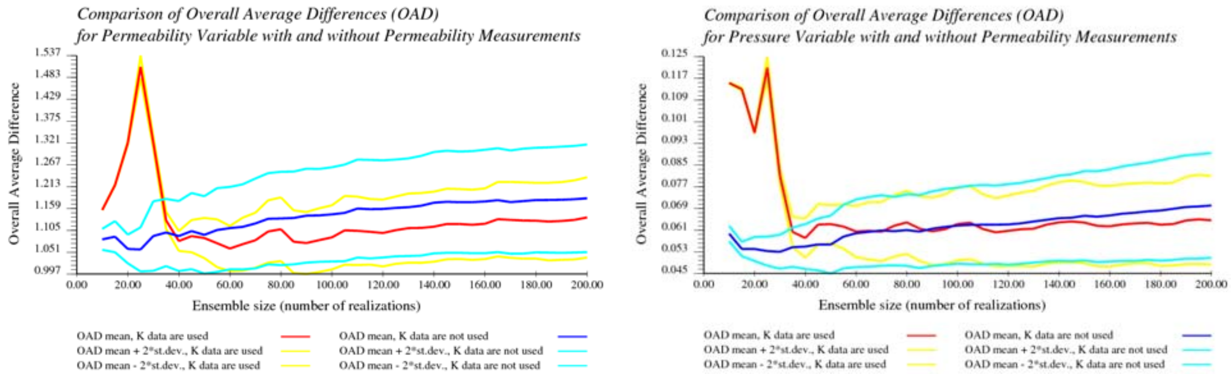

Figure 8 – Change of overall average differences (OAD) and two-standard-deviation confidence intervals with ensemble size of cases where permeability data are available and unavailable for permeability variable (left) and pressure (right) variables.

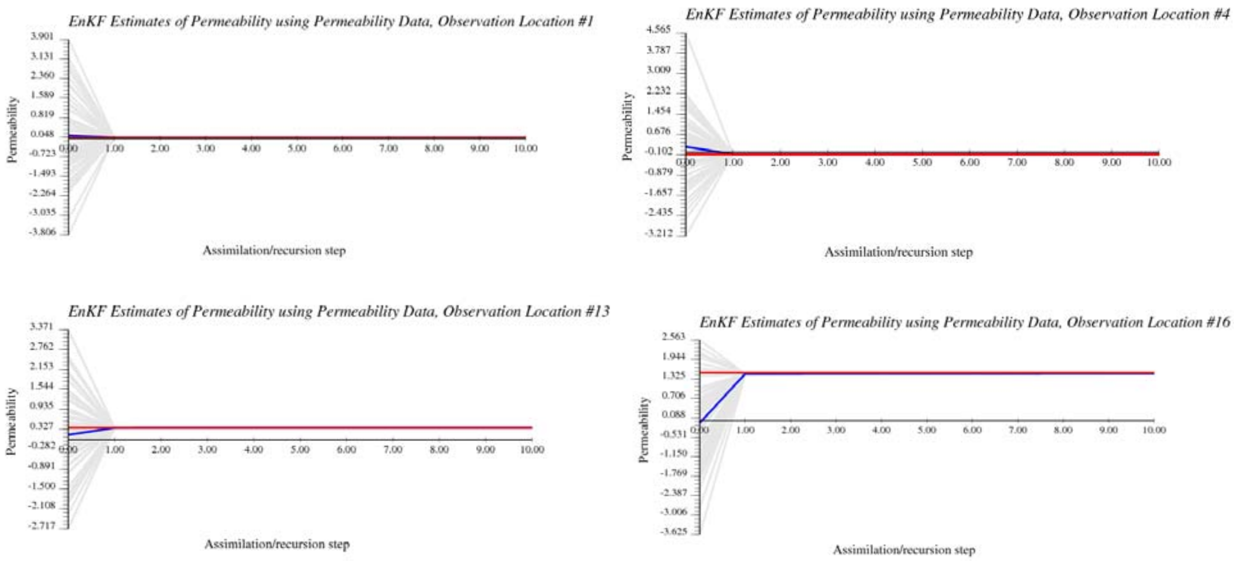

Figure 9 - Change of EnKF permeability estimates with assimilation step at observation locations #1, 4, 13, and 16. Ensemble of 60 realizations and permeability data are used. Value of single pressure realization is shown in grey, their average value – in blue, and true value – in red.

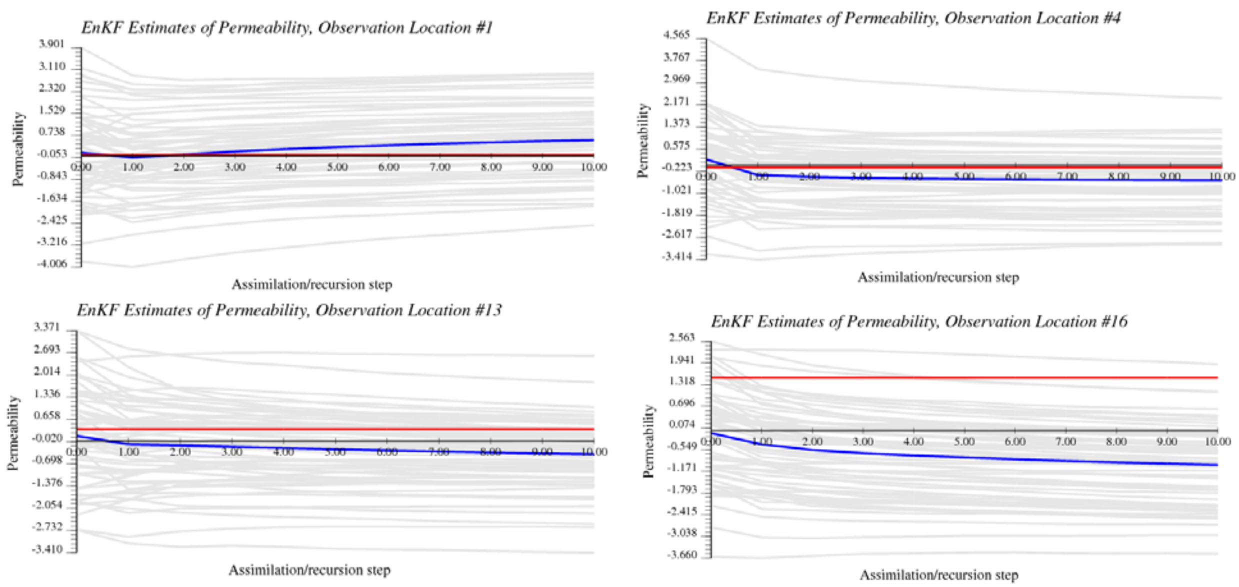

Figure 10 - Change of EnKF permeability estimates with assimilation step at observation locations #1, 4, 13, and 16. Ensemble of 60 realizations is used and permeability data are not available. Value of single pressure realization is shown in grey, their average value – in blue, and true value – in red.

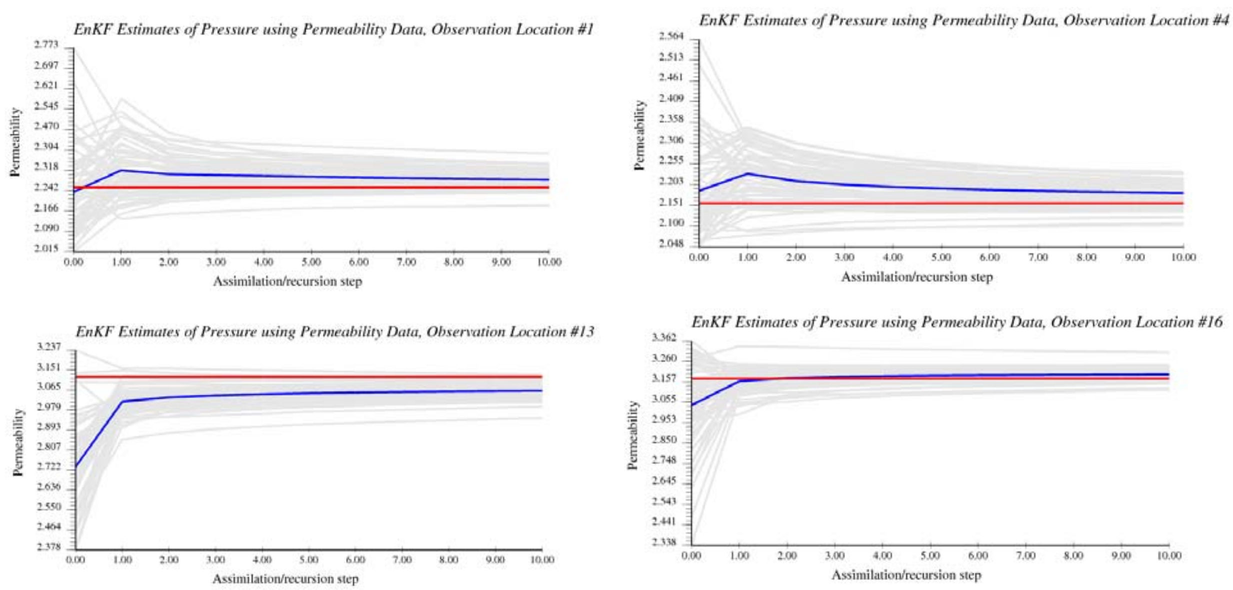

Figure 11 - Change of EnKF pressure estimates with assimilation step at observation locations #1, 4, 13, and 16. Ensemble of 60 realizations and permeability data are used. Value of single pressure realization is shown in grey, their average value – in blue, and true value – in red.

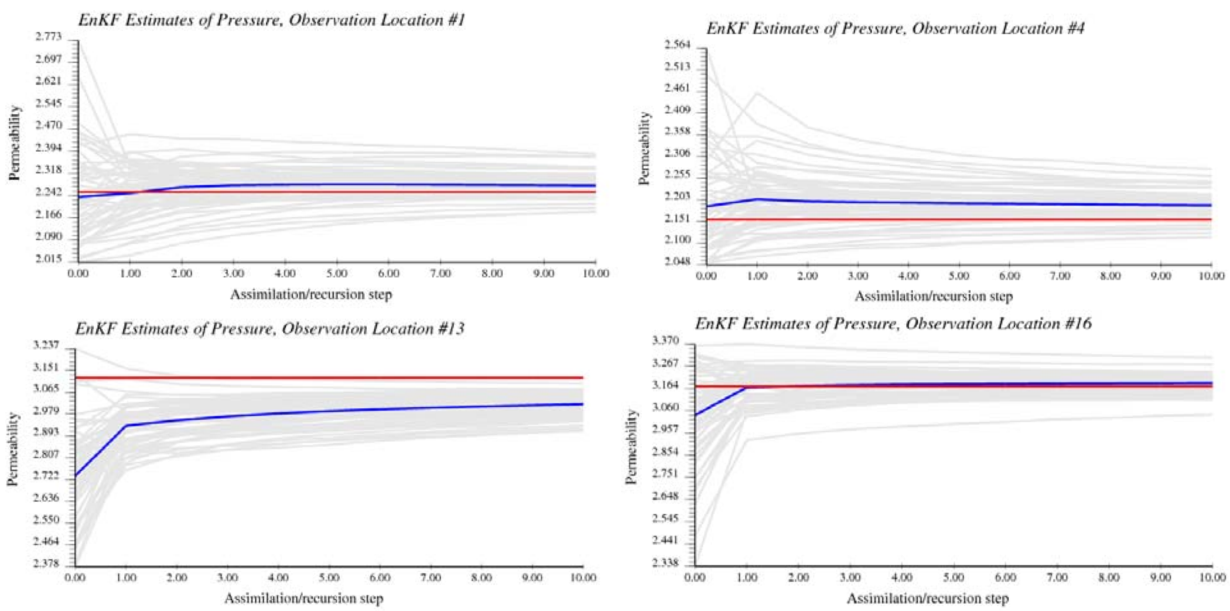

Figure 12 - Change of EnKF pressure estimates with assimilation step at observation locations #1, 4, 13, and 16. Ensemble of 60 realizations is used and permeability data are not available. Value of single pressure realization is shown in grey, their average value – in blue, and true value – in red.

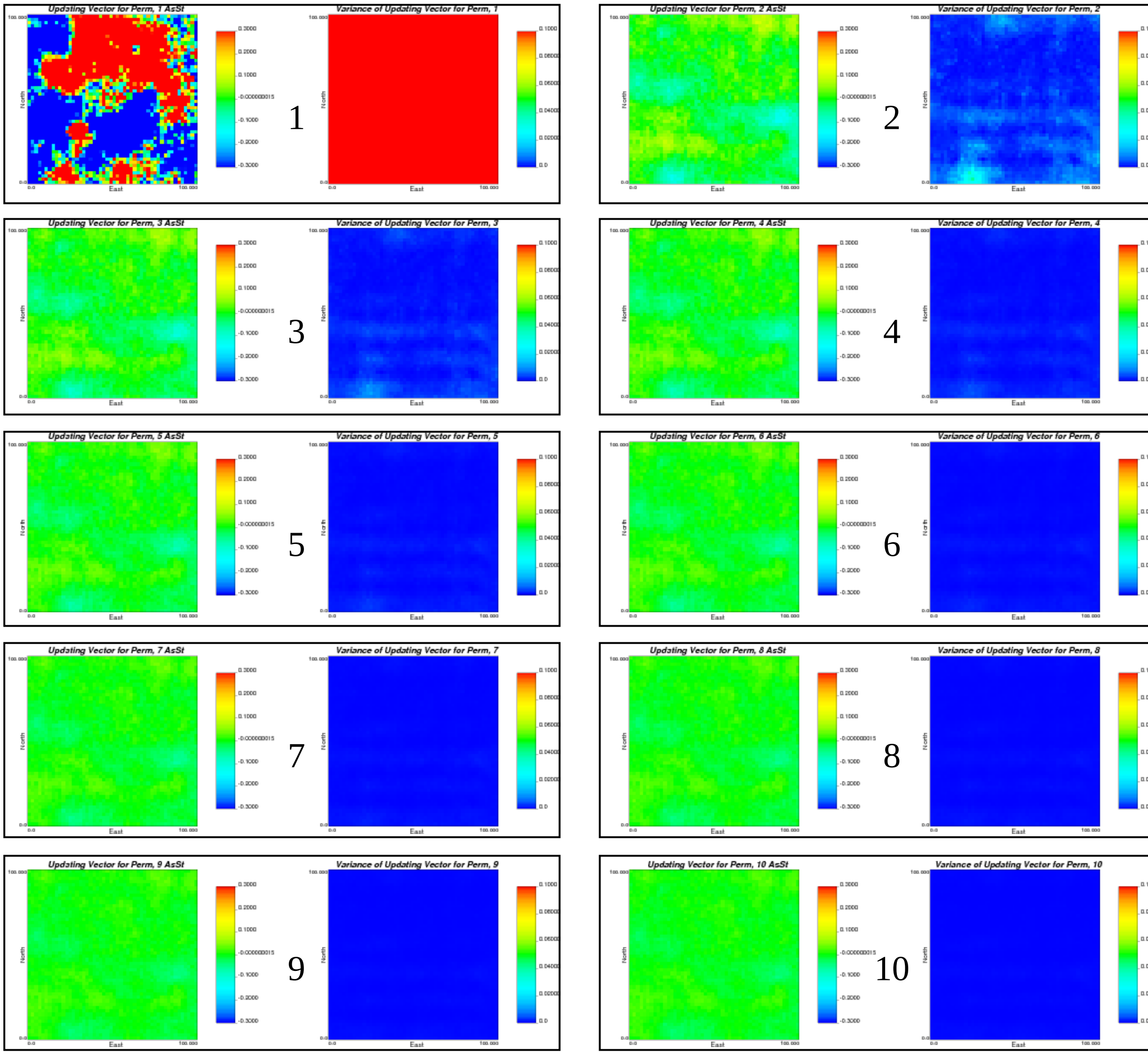

Figure 13 - Change of updating map of permeability and corresponding variance map with assimilation step, which is shown in the center of boxes. Range of scale bar of permeability updating vector is between -0.3 and 0.3, what is 10% of actual permeability scale, of variance – from 0.0 to 0.1. Ensemble of 60 realizations and permeability data are used.

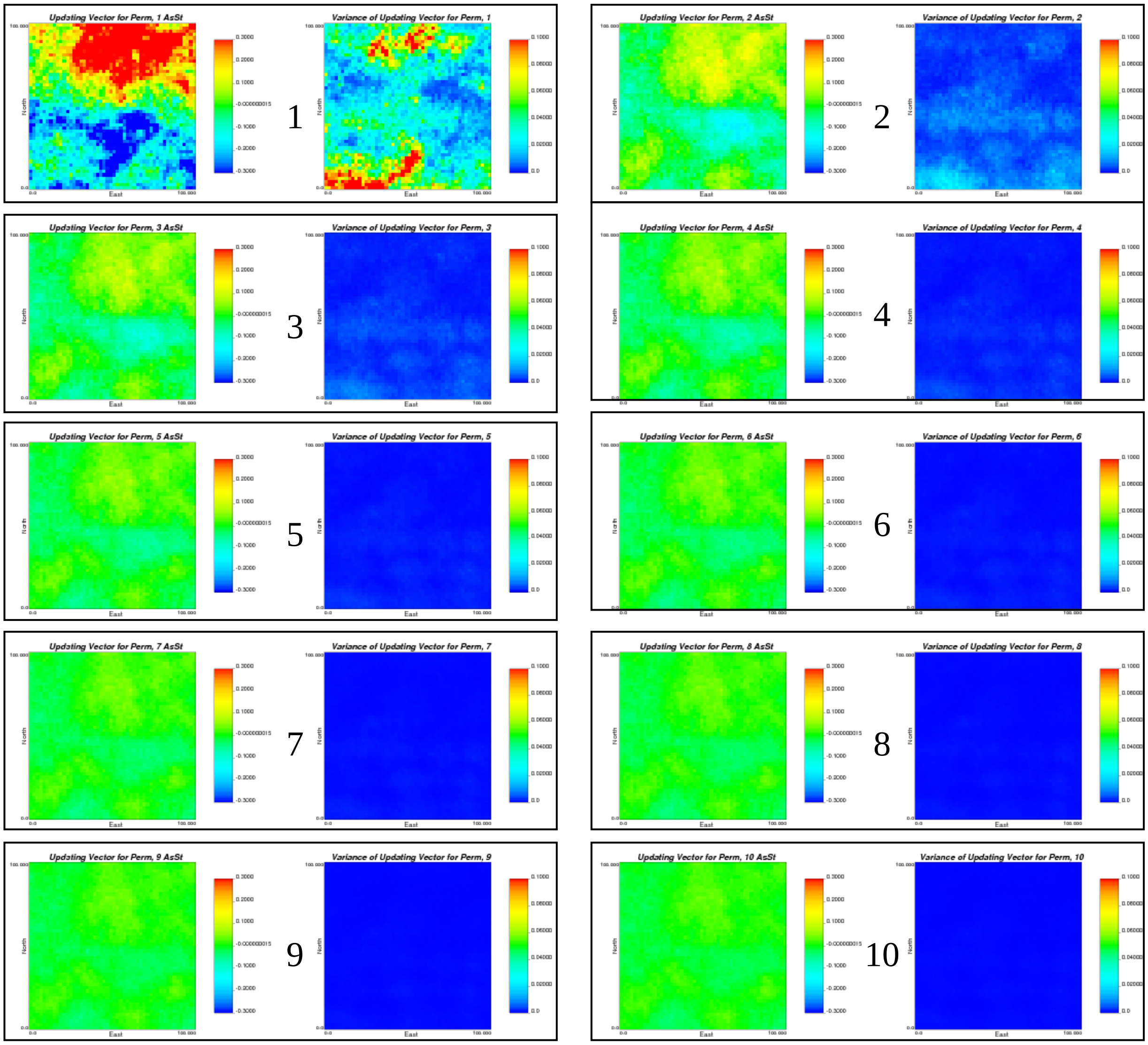

Figure 14 - Change of updating map of permeability and corresponding variance map with assimilation step, which is shown in the center of boxes. Range of scale bar of permeability updating vector is between -0.3 and 0.3, what is 10% of actual permeability scale, of variance – from 0.0 to 0.1. Ensemble of 60 realizations is used and permeability data are not available.

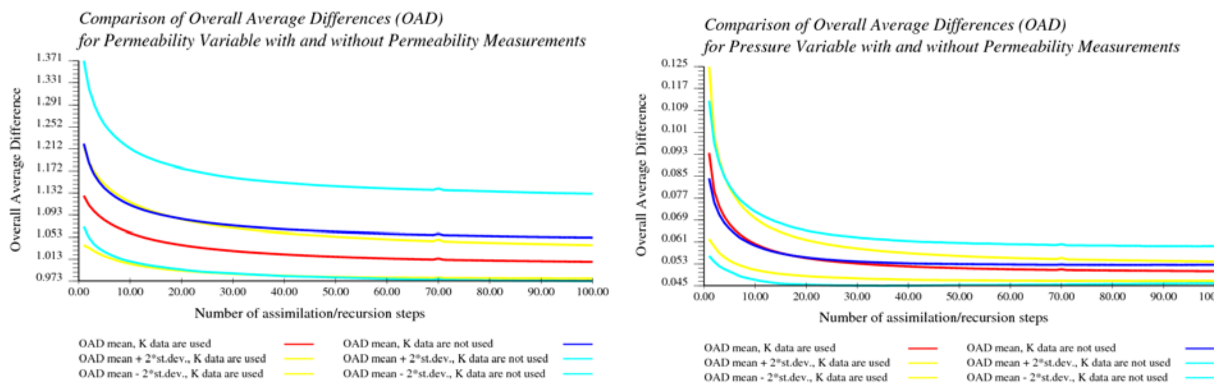

Figure 15 – Change of overall average differences (OAD) and two-standard-deviation confidence intervals with assimilation step of cases where permeability data are available and unavailable for permeability variable (left) and pressure (right) variables.

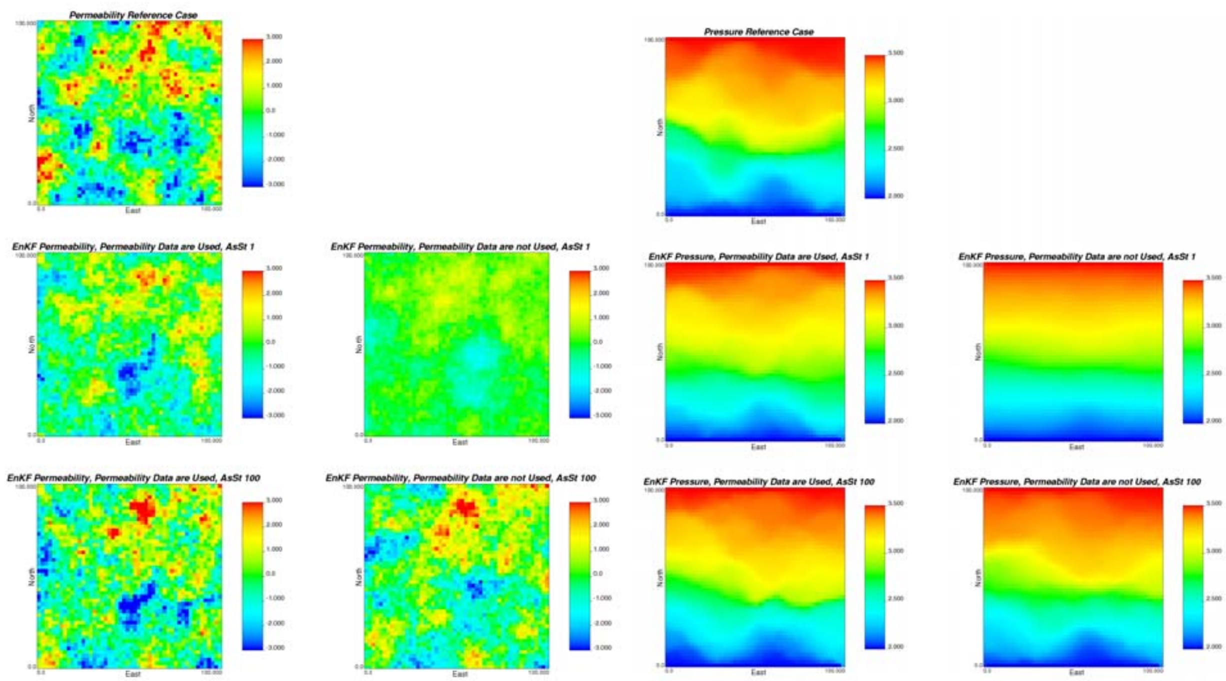

Figure 16 – Permeability and pressure reference cases and EnKF permeability and pressure estimates. Row 1 left – permeability reference case; row 2 left (from left to right) – E-type of EnKF permeability estimates using 60 realizations and permeability data after 1 assimilation step, E-type of EnKF permeability estimates using 60 realizations and no permeability data after 1 assimilation step; row 3 left (from left to right) E-type of EnKF permeability estimates using 60 realizations and permeability data after 100 assimilation steps, E-type of EnKF permeability estimates using 60 realizations and no permeability data after 100 assimilation steps; right part of rows is same as left, but for pressure.

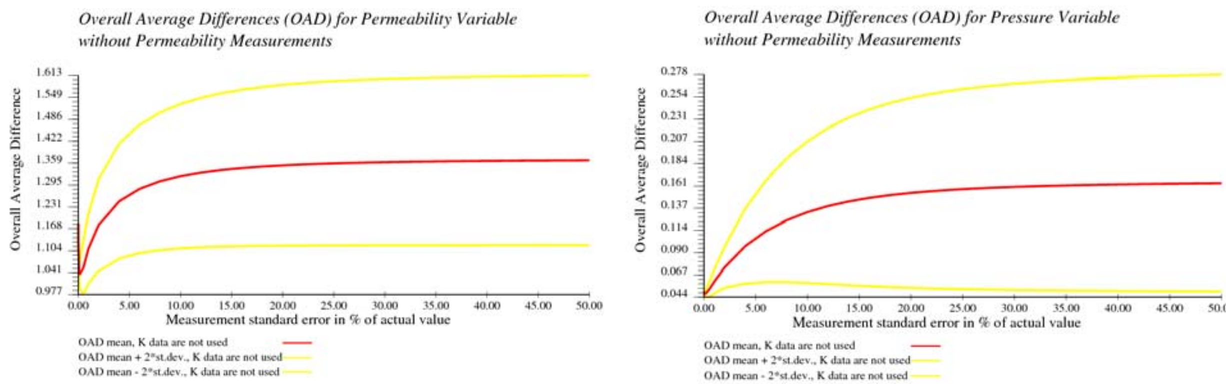

Figure 17 – Change of overall average differences (OAD) and two-standard-deviation confidence intervals with measurement error increase of case where permeability data are unavailable for permeability variable (left) and pressure (right) variables.

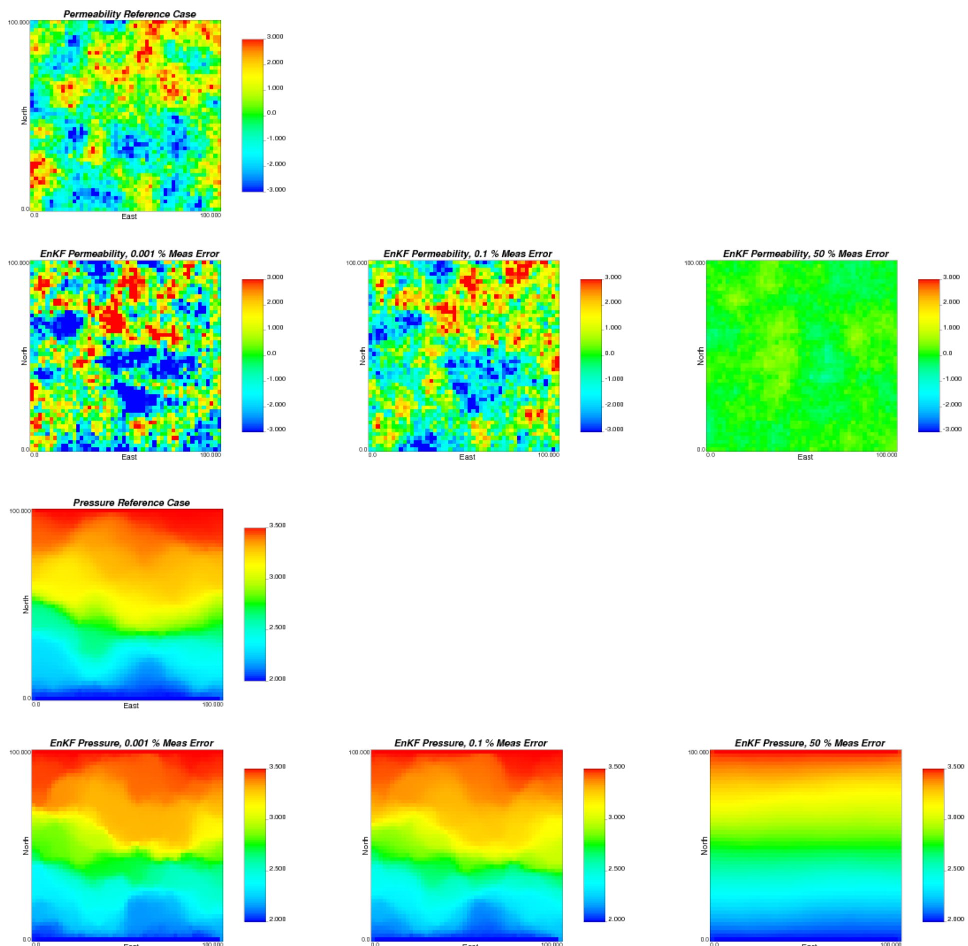

Figure 18 – Permeability and pressure reference cases and EnKF permeability and pressure estimates. Row 1 – permeability reference case; row 2 (from left to right) – E-type of EnKF permeability estimates using 60 realizations and no permeability data with 0.001% error in pressure data, E-type of EnKF permeability estimates using 60 realizations and no permeability data with 0.1% error in pressure data, E-type of EnKF permeability estimates using 60 realizations and no permeability data with 50.0% error in pressure data; row 3 is same as row 1, and row 4 is same as row 2, but for pressure variable.

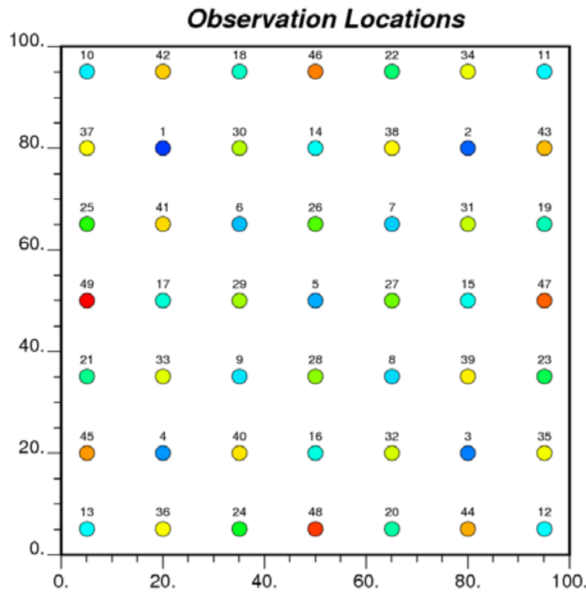

Figure 19 – Observation locations and order of data obtainment.

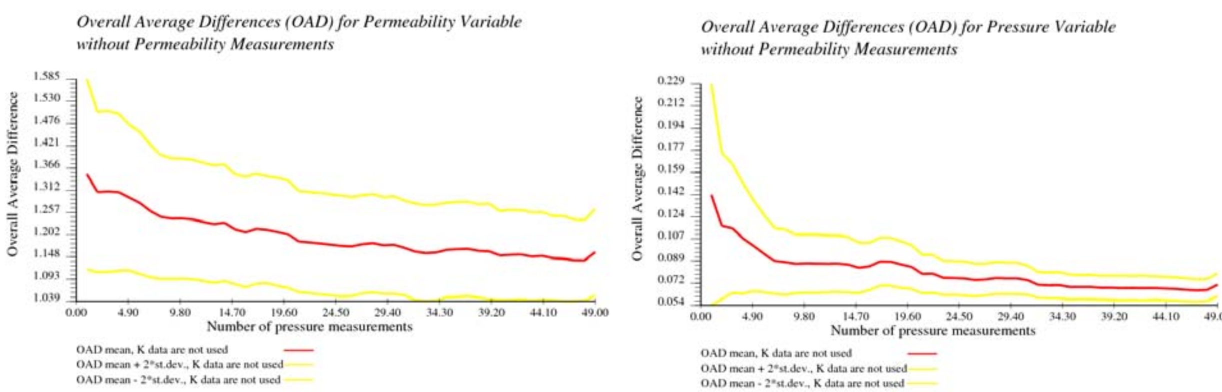

Figure 20 – Change of overall average differences (OAD) and two-standard-deviation confidence intervals with number of observations of case where permeability data are unavailable for permeability variable (left) and pressure (right) variables.

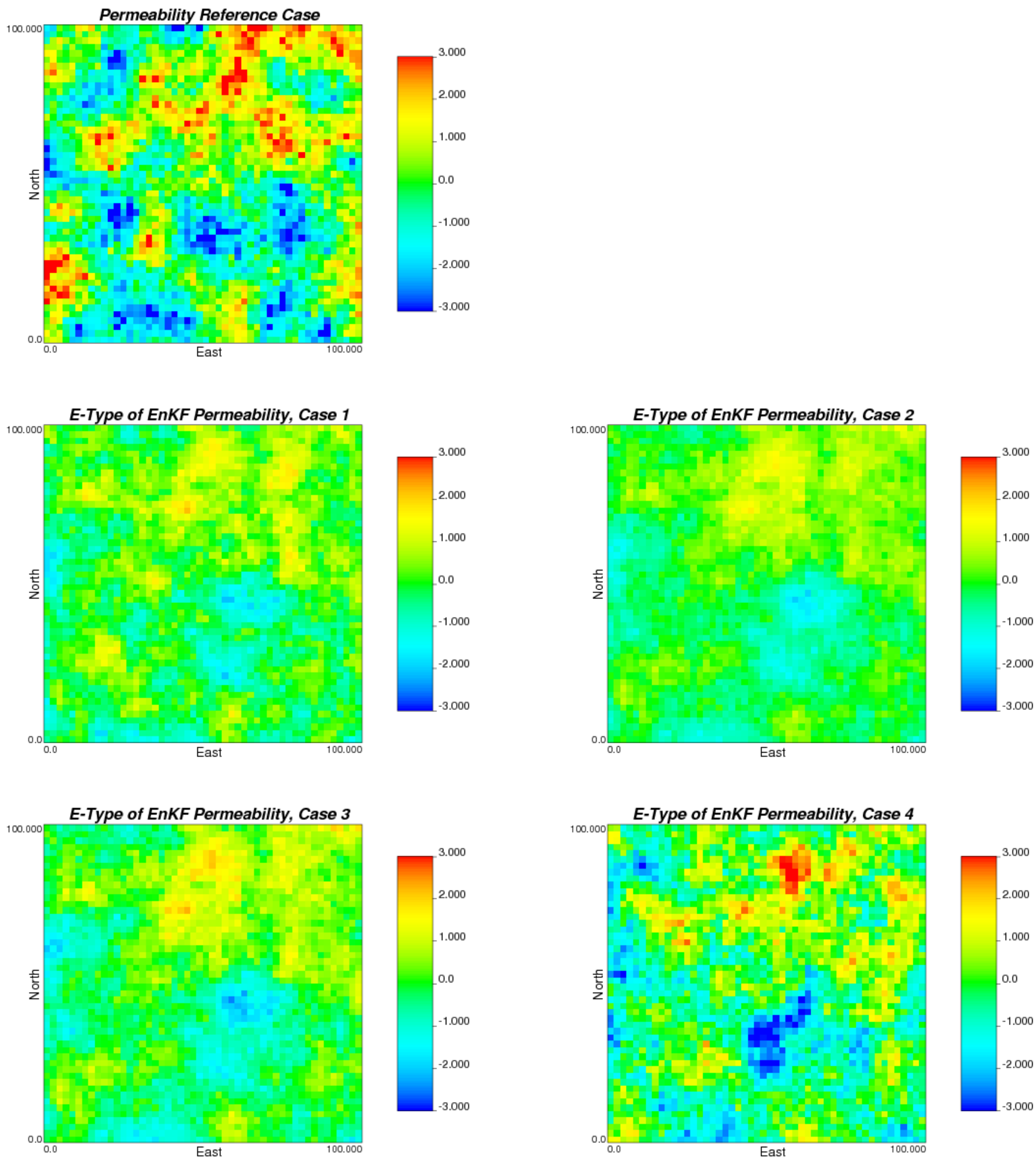

Figure 21 – E-type of permeability reference case (top) and of four cases. Case 1 – random initial permeability values, no permeability data; case 2 – initial permeability values and correct semivariogram model only, no permeability data; case 3 – initial permeability values and correct semivariogram model with same distribution range as reference case, no permeability data; case 4 – initial permeability values and correct semivariogram model with same distribution range as reference case and permeability data. 6o realizations are used.