Temperature and Seismic Data Integration with the Ensemble Kalman Filter (EnKF)

Yevgeniy V. Zagayevskiy and Clayton V. Deutsch

Vastly accumulated temperature data are continuously updated by measurements from permanent downhole gauges installed along vertical surveillance wells and thermally operated horizontal wells, but they are not widely used for characterization of geological properties of a petroleum reservoir model. Extensive 4D seismic attributes from repeated 3D seismic surveys conducted over same domain are frequently used only for qualitative description of reservoir performance. Ensemble Kalman Filter (EnKF) is proposed as a data assimilation technique for integration of temperature and seismic data to quantitatively describe architecture of a reservoir. Distribution of porosity and permeability fields is successfully constrained to hard and continuous soft data through EnKF on a small 2D synthetic SAGD case study. FORTRAN code `enkf.exe` *is written for this purpose. It is shown that incorporation of additional data acquired from different sources and time steps improves estimation of geological properties, and, thus, assists in prediction of flow zones and barriers. Implementation details of EnKF, remarks on integration of extensive dataset through localization techniques and shortcut for minimizing computational time are discussed as well.*

1. Introduction

Petroleum reservoir characterization is a continuous process and vital part of oil field development. Ultimate oil recovery and associated return heavily depend on quality of the reservoir model and especially on its geological component. Distribution of lithological facies determines flow zones, baffles and barriers, and, therefore, dictates flow paths of reservoir fluid and displacement agent. Facies model is populated with porosity and permeability values, which is used later on in flow simulation. Thereby, proper estimation of porosity and permeability fields is an essential task in geomodeling of a petroleum reservoir (Chen, 2007).

Geostatistical methods are usually employed to obtain best estimate of geological characteristics of a reservoir and to assess associated uncertainty. Data used in modeling may come from lab analysis of core, well tests, production observations and/or geophysical surveys and characterize either static or dynamic part of a model. The data themselves are divided into two broad classes: hard and soft data. While hard data are direct measurements of modeled variable, soft data are its indirect measurements, which are assimilated in a model to help in understanding of nature of the modeled variable. It is believed that incorporation of additional soft data can significantly improve knowledge about reservoir architecture. Seismic attributes and temperature measurements are good examples of soft data. 3D seismic surveys provide extensive spatial knowledge about reservoir geology. 4D seismic data, which are repeated in time measurements of 3D seismic response over fixed domain, may even be used to track fluid flow in a reservoir (Lumley, 2001), to predict front displacement of injected agent (Eastwood et al, 1994; Landrø, 2001; Yuh, 2003) or to forecast steam chamber growth in SAGD applications (Zhang et al, 2005). Steam assisted gravity drainage (SAGD) and cyclic steam stimulation (CSS) oil extraction techniques are widespread enhanced oil recovery techniques in Northern Alberta oil fields (Butler, 1991) that are home to enormous bitumen and heavy oil reserves. Steam is used as a displacement agent in these oil recovery processes, and, hence, is closely tied to spatial distribution of temperature profiles. Temperature data might be used in prediction of steam chamber growth and manage depletion process, since it is continuously measured by downhole permanent gauges installed in vertical surveillance wells and along horizontal well pairs.

Hard data for geological model, like porosity and permeability measurements, are mapped easily by means of geostatistical techniques, for instance, by kriging or sequential Gaussian simulation. However, assimilation of soft data requires more sophisticated modeling techniques, like kriging with external drift, Bayesian updating or simulated annealing (Deutsch, 2002). All these geostatistical techniques can model only one variable at a time using other variables as a secondary source. On the other hand, inverse modeling technique Ensemble Kalman Filter or simply EnKF, that utilizes concept of Bayesian inversion (Evensen, 2007), can be employed to integrate primary and secondary data in a petroleum reservoir model simultaneously. A methodology for integration of 4D seismic attributes and continuous reservoir temperature measurements to estimate porosity and permeability distributions using EnKF is proposed in this paper and implemented in FORTRAN program `enkf.exe`. While integration of seismic is not unique, temperature data assimilation is relatively novel idea.

Seismic data is a valuable source of geological information. However, integration of seismic attributes with EnKF is a challenging task, since highly extensive dataset may lead to collapse of ensemble of realizations that represent a model. Some alterations are introduced to EnKF in order to assimilate large datasets. Decomposition of covariance matrix (Skjervheim et al, 2005), hierarchical model (Myrseth, 2007), localization of updating matrix (Dong et al, 2006) or localization of covariance matrix (Fahimuddin, 2010) are implemented to continuously assimilate vast seismic data. Petroelastic model based on Gassmann's theory is used to generate synthetic seismic attributes for isothermal environments (Fahimuddin, 2010) and as well as for temperature dependent environments (Zagayevskiy and Deutsch, 2011).

Even though use of temperature data in estimation of various geological measures and flow rates may be found in the literature, this research topic is relatively new in petroleum reservoir geomodeling. Deconvolution with Bayesian inversion concept is applied by (Duru and Horne, 2009) to predict flow rate in conventional oil reservoir using temperature data from permanent downhole gauges. It is shown that temperature is a function of well product flow rate and thermal properties of fluid and rock surrounding a well, and, therefore, it is an additional constrain for reservoir geology. In subsequent paper of (Duru and Horne, 2010) temperature is utilized to estimate porosity and log-permeability distributions through quasilinear Bayesian inversion and EnKF, where flow and thermal models for conventional oil reservoir are coupled together. The authors argue that inclusion of temperature data improves estimation of geological properties in comparison to assimilation of only production data (bottom-hole pressure, oil and water production rates). Similar results are presented by (Li and Zhu, 2010; Li et al, 2011). Former paper touches a topic of prediction of water profile in horizontal well using bottom-hole pressure and temperature data measured by downhole gauges and fibre optic sensors. In second paper authors show that incorporation of production data and bottom-hole pressure with temperature data to geological model helps predicting permeability distribution and leads to optimized control of horizontal well performance. To be more specifically, nonisothermal reservoir model and temperature data are used to derive coarse-scale permeability field, which is downscaled by block kriging and then calibrated to permeability distribution derived from production data. Comparison study is conducted as well. Integration of temperature data by means of EnKF is compared to sequential self-calibration technique (SSC) (Gómez-Hernánez et al, 1997), where rate of minimization of objective function is faster for EnKF (Zagayevskiy and Deutsch, 2010). Also integrated modeling approaches, in which beside temperature and production data other soft data are integrated in a reservoir model, exist. For instance, temperature, hydrologic, pneumatic and geochemical data are used to characterize flow and transport mechanisms happening in fractured reservoirs (Wu et al, 2007).

Attempts to use seismic and temperature data simultaneously to constrain reservoir model are made as well. Temperature dependent seismic velocities from 4D seismic surveys are employed to predict steam flood distribution in 5-spot pattern at South Casper Creek oil field in Wyoming (Johnston et al, 1992).

This paper is organized as follows. First, theoretical background of EnKF is presented. Second, the proposed methodology for integration of temperature and pressure dependent seismic attributes and temperature data for estimation of porosity and permeability distributions is thoroughly explained. Third, implementation details of data assimilation by means of EnKF, remarks on integration of large datasets through localization and shortcut for reduction of computational cost are discussed and shown on small examples. Four, description of parameter file of FORTRAN program `enkf.exe` is presented. Finally, synthetic 2D case study for integration of time-lapse seismic and continuously measured temperature data from surveillance wells for SAGD well pair is shown and conclusion is made.

2. Theoretical Background of EnKF

Data integration is an optimization process that implies imposing constrains to a numerical model in order to make it plausible, honor all available data and minimize objective function. A procedure of minimization of objective function is called a history match in petroleum reservoir engineering. Objective function is usually expressed as a weighted sum of squared difference between estimated and true values (data) of a model (see the Equation (1)).

$$O = \sum_{t=1}^{N_t} \sum_{i=1}^{N_{Dt}} w_{i,t} \cdot \left(Z_{i,t} - D_{i,t}\right)^2 \qquad (1)$$

where, O is the objective function, which has to be minimized; w_i is the weight of i^{th} datum; N_t is the number of time steps from which data are coming; N_{Dt} is the number of data assimilated at time t; $Z_{i,t}$ and $D_{i,t}$ are the i^{th} model and associated datum values at time t.

Data integration may be performed once to build a static model, like in interpolation methods, or continuously carried out to build a dynamic model as new observation becomes available. A dynamic model consists of model parameters, which are constant in time, and state variables, which change in time and are function of model parameters. Both variable types are constrained to assimilated data through minimization of objective function and system of governing equations. Mathematical expression of a dynamic model is shown in the Equation (2). Petroleum reservoir model is a good example of dynamic model that is governed by mass and momentum balances and energy conservation equations. Estimate does not necessarily match truth exactly, and, thus, more availability of information leads to a higher quality model.

$$\boldsymbol{Z}_t = \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Y}_t \\ \boldsymbol{D}_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{X} \\ f(\boldsymbol{X},t) \\ g(\boldsymbol{X},t) \end{bmatrix} \tag{2}$$

where, $\boldsymbol{Z}_t$ is the vector representation of a dynamic model at time t; $\boldsymbol{X}$ is the vector of models parameters that do not vary in time; $\boldsymbol{Y}_t$ is the vector of state variables at time t; $\boldsymbol{D}_t$ is the data vector integrated in a model at time t; f and g are the functions that relate model parameters to state variables and data respectively.

Ensemble Kalman Filter (EnKF) is an inverse modeling technique based on Bayesian approach. Set of statistically probable realizations represents a dynamic model, where data are continuously integrated in the model honoring all previously assimilated data and minimizing objective function. EnKF estimates model parameters and predicts state variables through data assimilation, linear covariance function between the variables and governing equations. Ensemble Kalman Filter has evolved from Kalman filter. In EnKF covariance matrix is replaced with sample covariance matrix derived from ensemble members. The EnKF consists of two steps: nonlinear forecast (propagation in time or prediction step) and linear analysis (model update step through data integration). These two steps are jointly called an assimilation step. The equations of EnKF are shown below. More information on the filter can be found in (Evensen, 2007; Zagayevskiy et al, 2010a and 2010b). Note that in EnKF all variables $\boldsymbol{X}$, $\boldsymbol{Y_t}$, $\boldsymbol{Z_t}$, $\boldsymbol{D_t}$ are presented in matrix form of size N_X x N_e, N_Y x N_e, N_{Dt} x N_e and $(N_X + N_Y + N_{Dt})$ x N_e respectively, where N_X and N_Y are the number of model parameters and state variables; N_{Dt} is the number of data assimilated at time t holding inequality $N_{Dt} \leq N_X + N_Y$; N_e is the ensemble size or number of realizations. It is recommended to keep ratio of number of data to ensemble size around 1:5 (Fahimuddin, 2010). EnKF is good for estimation and prediction of large linear systems that follow normal or Gaussian distribution.

Forecast step equation:

$$\boldsymbol{Z}_t^f = M\left(\boldsymbol{Z}_{t-1}^a\right) + \boldsymbol{E}_{t-1}^{model} \tag{3}$$

Analysis step equation:

$$\boldsymbol{Z}_t^a = \boldsymbol{Z}_t^f + \boldsymbol{K}_t \cdot \left(\boldsymbol{D}_t - \boldsymbol{H}_t \cdot \boldsymbol{Z}_t^a\right) \tag{4}$$

Kalman gain equation:

$$\boldsymbol{K}_t = \hat{\boldsymbol{C}}_t^f \cdot \boldsymbol{H}_t^T \cdot \left(\boldsymbol{H}_t^T \cdot \hat{\boldsymbol{C}}_t^f \cdot \boldsymbol{H}_t^T + \boldsymbol{R}_t\right)^{-1} \tag{5}$$

Sample covariance matrix equation:

$$\hat{\boldsymbol{C}}_t^f = \frac{\left(\boldsymbol{Z}_t^f - \overline{\boldsymbol{Z}}_t^f\right) \cdot \left(\boldsymbol{Z}_t^f - \overline{\boldsymbol{Z}}_t^f\right)^T}{N_e - 1} \tag{6}$$

where, ^{f} and ^{a} superscripts that represent model at forecast and analysis steps respectively; *M* is the model operator that relates model *Z* values from different adjacent time steps, while model parameters do not change at this step, state variables are propagated in time; $E^{model}{}_{t}$ is the error in the model at time *t*, it is assumed to be zero in this paper; K_t is the Kalman gain at time *t*, which stores linear covariance function of model variables; H_t is the observational matrix that relates model parameters or state variables to appropriate data; $\hat{\boldsymbol{C}}_t^f$ is the sample covariance matrix at time *t* between all forecasted model variables computed from realizations; $\boldsymbol{R}_t$ is the measurement error matrix of the data at time *t*; $\overline{\boldsymbol{Z}}_t^f$ is the mean of model values over realizations.

It might be noticed that analysis step of EnKF is similar to kriging linear estimate in mathematical aspect. This fact has been proven before (Zagayevskiy et al, 2010a). Here it is shown that estimation results from sequential Gaussian simulation (SGS) and analysis step of EnKF are exactly the same: compare mean and estimation variance derived from both techniques (Figure 9, estimation variance is not shown). Moreover, it is clear that sequential assimilation of data honor all previously integrated data and final estimate is exactly the same as if data assimilated altogether at once. This feature is vivid advantage of EnKF over SGS. This observation is obvious outcome of mathematical framework of EnKF.

Algorithm of Ensemble Kalman Filter can be described as follows:

1. Generate initial ensemble $\boldsymbol{X}^{a}_{t=0}$ of model parameters (usually it is geological properties of a model) using all prior information about the model and run first forecast step (Equation (3)) to get initial ensemble of state variables $\boldsymbol{Y}^{a}_{t=1}$. Thus, full model $\boldsymbol{Z}^{f}_{t=1}$ is defined at first forecast step.
2. Execute analysis step (Equation (4)), where entire model is updated honoring all available data $\boldsymbol{D}_{t=1}$ at certain time step *t*.
3. Go back to step #1 using updated values as initial ensemble to forecast state variables at next time step $t = t + 1$ and update ensemble with newly acquired data as it is performed in the step #2. Continue coupled process until all data from various time steps are integrated in the model. Last forecasted ensemble can be used for prediction of state variables.

3. Methodology for Integration of Seismic Attributes and Temperature Data

A methodology for integration of seismic attributes and temperature data to a reservoir model by means of EnKF can be summarized in a chart shown in the Figure 1. The objective is to characterize geology of heavy oil and bitumen reservoirs (estimate porosity and permeability values) and to help managing of thermally operated production processes (predict steam chamber growth in SAGD oil recovery methods). For this reason EnKF is proposed as modeling technique and is treated as black box, i.e. only input and output of a reservoir model are important and internal mechanism may be unknown to some extent. The methodology is closely related to the algorithm described in previous section.

Proposed reservoir model *M* can be divided into three sub models: geological, production, and petroelastic components, which are related to each other in well-established manner. Geological part of a reservoir model represents model parameters and consists of porosity scalar φ and permeability tensor $\boldsymbol{K}$, which is assumed to be a diagonal matrix. Permeability values in all three directions may be expressed through values of one of tensor element as shown in the Equation (7), where α and β are operators or coefficients that relates diagonal elements to each other. In turn, logarithmic values (base 10) of permeability in one direction may be obtained from porosity and regression model (Deutsch, 2002). The simplest form of regression model is a linear one, which is shown in the Equation (8). Program `poroperm.exe` is used to generate synthetic permeability values based on porosity values using equation of fitted regression line, linear correlation coefficient and variance of porosity computed from the data. Spatial correlation between variables is not assumed in this model, only bivariate relationship is reproduced. While porosity distribution is usually normal, generated permeability values follow lognormal distribution. Description of the program and its theoretical background are given in the Appendix.

$$\begin{aligned} K_{xx} &= K_{xx} \\ K_{yy} &= \alpha(K_{xx}) \\ K_{zz} &= \beta(K_{xx}) \end{aligned} \qquad (7)$$

$$\log(K_{xx}) = a + b \cdot \varphi + \varepsilon \quad (8)$$

where, K_{xx}, K_{yy}, and K_{zz} are the diagonal components of permeability tensor ***K***, all other non-diagonal components are assumed to be zero; a and b are the regression coefficients of linear model between logarithmic permeability and porosity derived from the data; ε is the error in regression model, which introduces probabilistic feature to linear regression model.

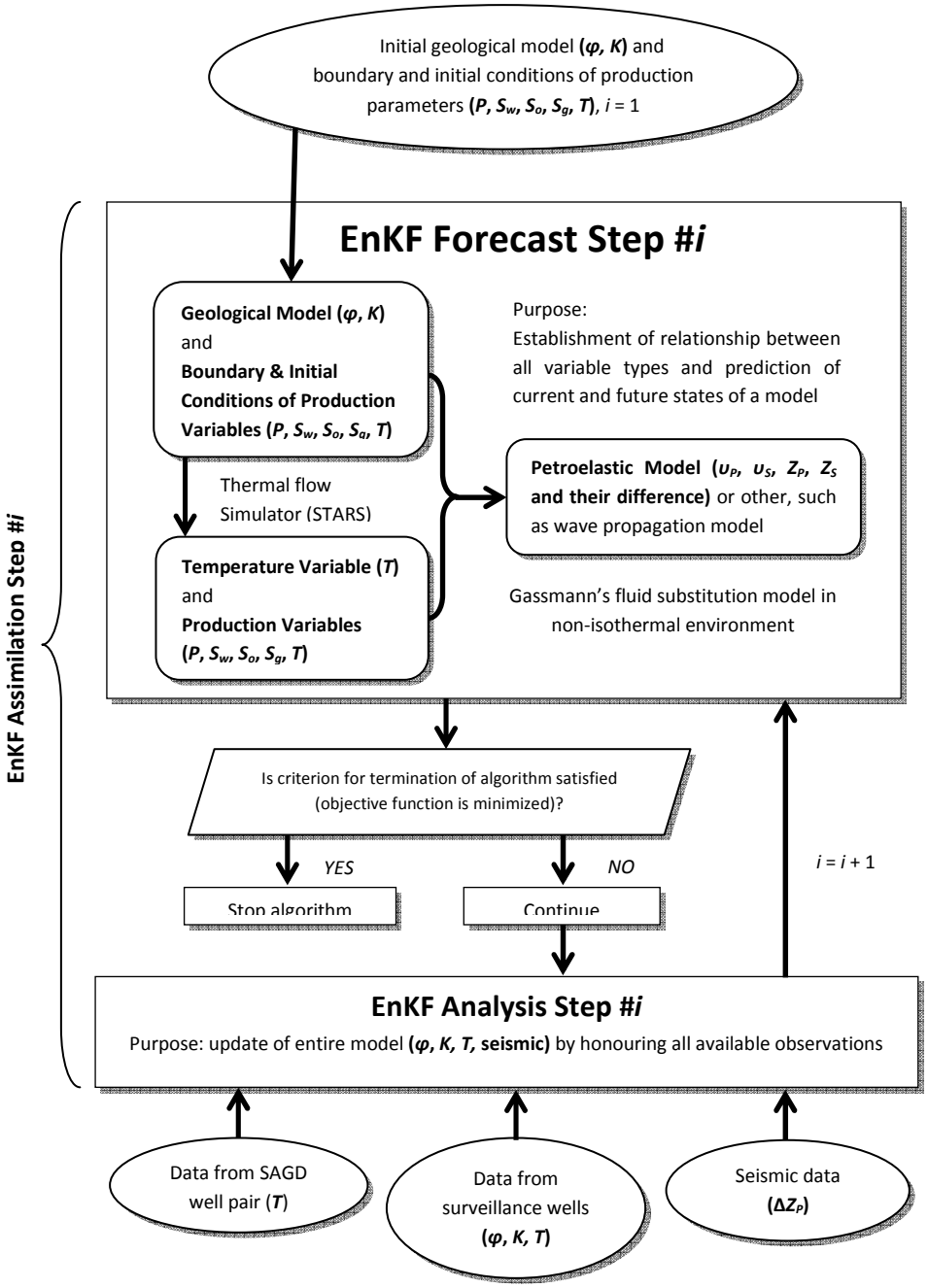


Figure 1: Schematic of methodology for integration of seismic and temperature data

Production and petroelastic sub models represent state variables of a reservoir model, which vary in time. In this methodology main production state variables are reservoir pressure P and temperature T, water S_w, oil S_o,

and gas S_g saturations. They are simulated using geological properties of a reservoir and thermal flow simulator, like CMG`s STARS.

Petroelastic model generates synthetic seismic attributes. Gassmann`s fluid substitution model is selected as basis for modeling seismic velocities $V_p$, $V_s$ and associated acoustic impedances $Z_p$, $Z_s$ (Gassmann, 1951). Other models could be selected as well, for instance, wave propagation model. Here Gassmann`s model depends on reservoir pressure and temperature in order to preserve relationship between temperature and seismic attributes propagated in nonisothermal environment of thermally operated oil fields. Theoretical background of the petroelastic model can be found in (Batzle and Wang, 1992; Zagayevskiy and Deutsch, 2011). Generated seismic attributes are function of porosity, fluid saturations, mineralogical content of rock and their physical and elastic properties.

Methodology for seismic and temperature data assimilation through EnKF can be described as follows (Figure 1). In order to initiate EnKF algorithm initial ensemble of porosity is required. Sequential Gaussian simulation is used to generate initial porosity values. Spatial distribution and all prior information should be embedded into the initial ensemble. Initial permeability ensemble is derived from porosity-permeability transform (Equations (7) and (8)). Once geological properties are defined, flow simulator is run to get initial ensemble of production state variables, which are mainly pressure, temperature and fluid saturations, at time *t*. Output of geological and production sub models are used in petroelastic sub model to generate pressure and temperature synthetic seismic attributes and/or their difference. It has been shown that the attributes are good to reveal geological properties, but difference between seismic attributes from baseline and subsequent surveys help to track propagation of steam flooding zones or steam chamber growth (Zagayevskiy and Deutsch, 2011). At this point entire reservoir model is initialized. Criterion for goodness of model estimation is checked against specified value. Objective function (Equation (1)) can be used as a criterion to move to analysis step at time *t* or terminate EnKF. At analysis step all available data observed at time *t* are integrated into a model using the Equation (4). Thus, the process is repeated until no more newly acquired data are available or termination criterion is satisfied.

4. Implementation Details

Some implementation details and concerns regarding integration of temperature data and extensive seismic attributes in a petroleum reservoir model through EnKF are discussed here. They are briefly summarized below:

- Initial ensemble
- Ensemble size
- Units of variables
- Computational time

4.1 Initial Ensemble

Generation of initial ensemble is a vital procedure, since stored covariance structure inside it influences estimation of model parameters and prediction of state variables at all subsequent time steps. It is recommended to generate initial ensemble of model parameters following normal distribution first, and then derive initial ensemble of state variables using model operator (flow simulator and petroelastic model in our case) and initial values of model parameters. Values from different realizations have to be spatially correlated to each other. SGS with proper semivariogram model $\gamma(\boldsymbol{h})$ derived from available data is one of the best ways for generation of initial ensemble (Deutsch, 1998). If generated initial or any other ensembles have distribution that is too far from Gaussian, normal score transformation may be applied to a variable of interest. Also patterns of spatial continuity of variables stored in the ensemble should depict realistic configuration. Otherwise very weak estimate is obtained. See more details in section on computational time.

4.2 *Ensemble Size*

It is claimed that ensemble size should be at least ten times greater than number of assimilated data (Fahimuddin, 2010), otherwise spurious long-range covariance matrix is obtained that leads to collapse of the ensemble. The reason lies in insufficient number of degrees of freedom and loss of matrix rank. Therefore, if extensive dataset, such as 4D seismic attributes, is integrated in a model, very large number of realizations is required to get reliable estimate, what causes increase in computational time. In order to solve this issue localization of updating matrix or localization of covariance matrix may be applied. Localization leads to a local update of the ensemble matrix with

possible artifacts depending on updating window size or selected smoothing function, and, thus, it is an approximation of global update with worse estimation accuracy.

Localization of updating matrix implies that only ensemble members lying in a region around assimilated data are updated at analysis step. Rest of the ensemble is kept unchanged. The updating region is specified by a user as an updating window and is intended for elimination of long-range spurious correlation between model realizations. Mathematically local update based on specified window size can be expressed as shown in the Equation (9). One datum at time assimilation is implemented in program `enkf.exe` using only localization of updating matrix with rectangular updating window, whose size is specified by a user. An example of localization of updating matrix is shown in the Figure 10. Base case of 1500 grid block 2D model (50 blocks in X direction and 30 in Z direction) represents a difference in acoustic impedances from baseline and subsequent surveys. 250 data are integrated in the petroelastic model to restore full distribution of the variable. Three different cases are examined: global EnKF update with 1000 realizations, global EnKF update with 100 realizations and local EnKF update with 100 realizations and window size of 11 x 1 x 11 blocks. It is obvious that global best estimate of 1000 realizations matches base case pretty well. Global best estimate of 100 realizations is good representative of collapsed ensemble matrix, which is caused by dominance of spurious long-range covariance structure and insufficient degrees of freedom. On the other hand localization of updating matrix, and, therefore, avoiding influence of spurious long-range covariance, leads to good estimate using the same 100 realizations. Note that even though means from global and local EnKF approaches are close to each other, they are not exactly the same, what proves the fact that local update is just approximation of global one.

$$\begin{aligned} \boldsymbol{Z}_t^{a,i+1} = \boldsymbol{Z}_t^{a,i} + \hat{\boldsymbol{C}}_t^{f,i} \cdot \boldsymbol{H}_t^T \cdot \left\{\boldsymbol{L}_t^{u,i}\right\}^T \cdot \left(\boldsymbol{L}_t^{u,i} \cdot \boldsymbol{H}_t \cdot \hat{\boldsymbol{C}}_t^{f} \cdot \boldsymbol{H}_t^T \cdot \left\{\boldsymbol{L}_t^{u,i}\right\}^T + \boldsymbol{L}_t^{u,i} \cdot \boldsymbol{R}_t \cdot \left\{\boldsymbol{L}_t^{u,i}\right\}^T\right)^{-1} \\ \cdot \left(\boldsymbol{L}_t^{u,i} \cdot \boldsymbol{D}_t - \boldsymbol{L}_t^{u,i} \cdot \boldsymbol{H}_t \cdot \boldsymbol{Z}_t^{a}\right), \\ i = 0,..,n_{lu} \quad \text{and} \quad \boldsymbol{Z}_t^{a,0} = \boldsymbol{Z}_t^{f}, \quad \boldsymbol{Z}_t^{a,1+n_{lu}} = \boldsymbol{Z}_t^{a} \end{aligned} \tag{9}$$

where, $\boldsymbol{L}^{u,i}{}_t$ is the updating window matrix of size $N_{Dt,I}$ x N_{Dt} consisting of zeros and ones and determines size of updating window for ensemble analysis and which members are updated; $N_{Dt,i}$ is the number of data falling into a an updating window; n_{lu} is the number of times local update is applied at analysis step at assimilation time step *t* with index *i*.

Localization of covariance matrix is slightly different from localization of updating matrix. It has higher accuracy, but it is more complicated in implementation. Localization of covariance can be expressed as shown in the Equation (10), where only Kalman gain is modified keeping other EnKF equations unchanged. Symbol ◦ stands for element-wise product of two matrices of same size (Equation (11)). Localization of covariance matrix is implemented through smoothing positive definite matrix $\boldsymbol{L}^{c}{}_t$ eliminating influence of long-range relationship at analysis step, and, thus, preserving rank of matrix. Fifth-order correlation function of Gaspari and Cohen is widely used as covariance localization matrix $\boldsymbol{L}^{c}{}_t$ (Fahimuddin, 2010). No example is provided.

$$\boldsymbol{K}_t = \left\{\boldsymbol{L}_t^c \circ \hat{\boldsymbol{C}}_t^f\right\} \cdot \boldsymbol{H}_t^T \cdot \left(\boldsymbol{H}_t^T \cdot \left\{\boldsymbol{L}_t^c \circ \hat{\boldsymbol{C}}_t^f\right\} \cdot \boldsymbol{H}_t^T + \boldsymbol{R}_t\right)^{-1} \tag{10}$$

$$\left\{\boldsymbol{L}_t^c \circ \hat{\boldsymbol{C}}_t^f\right\}_{i,j} = \left(\boldsymbol{L}_t^c\right)_{i,j} \cdot \left(\hat{\boldsymbol{C}}_t^f\right)_{i,j} \tag{11}$$

where, $\boldsymbol{L}^{c}{}_t$ is the covariance localization matrix of size $(N_X + N_Y)$ x N_e; ◦ stands for element-wise matrix product or Hadamard product.

4.3 *Units of Variables*

There is no need to modify or standardize in any way units of variables in EnKF. Forecast step requires input values in original units and generates output variables in original units as well. Kalman gain in analysis step is the only component that might be sensitive to choice of units. However, because of its nature (see the Equation (5)) Kalman gain is a dimensionless matrix, and, thus, it does not depend on units. However, in some cases, as described in section on generation of initial ensemble, variable's transformation to Gaussian space may be

required, if original distribution is too far from normal. Example of temperature field, which does not follow normal distribution, is shown in the Figure 11 and Figure 12. 32 data are integrated in a model, where analysis step is carried out either in original or normal scored units. It is obvious that best estimate derived using normal score transformation is much closer to the base case than EnKF mean derived from use of original units.

Normal score transformation should not be performed, when limited number of data is present or numerous data of exactly same values are integrated in a model, since back transformation does not lead to a unique solution. Comparison of porosity data integration with and without normal score transformation is presented in the Figure 13. Here porosity represents discrete variable, which can take only two values: high for reservoir and low for baffle zones. Back transformation of porosity estimates from Gaussian space to original units produces artifacts, which are not seen, when analysis step is carried out in original units.

4.4 Computational Time

Two main factors determine computational cost of EnKF beside the model size. They are ensemble size (number of realizations) and number of times flow simulator is run at forecast step. Latter one comprises larger portion of computational overburden. Attempts are made to decrease computational time and increase estimation quality of the model. As it was mentioned before one of possible and effective solutions to decrease number of realizations is to apply localization of updating or covariance matrices. The localization reduces required size of ensemble matrix and saves time at analysis step. However, it should be noted that in localization scheme decrease of computational time is not always proportional to decrease of number of realizations, since iterative procedure may be required as shown in the Equation (9). Thus, number of integrated data is also important factor of computational cost of EnKF analysis step.

In order to minimize number of times flow simulator is run the shortcut based on concept of mean of realizations is proposed and described in details below. While conventional forecast step is depicted schematically in the Figure 2, proposed shortcut is shown in the Figure 3. In conventional procedure flow simulator is run for every single realization of model parameters to obtain corresponding realization of state variables. When ensemble size is large and a reservoir model is not simple, application of EnKF to data integration becomes impractical due to very high computational expenses. If shortcut based on mean of realizations concept is applied, flow simulator should be run only once, what can save time considerably. The proposed methodology is as follows. Mean of model parameters is used to get mean of state variables through running flow simulator once. N_e realizations of spatially correlated random numbers are simulated by means of sequential Gaussian simulation unconditional to data, but conditional to model parameters and with proper semivariogram model obtained from data on state variables. Thus, generated SGS realizations should have zero mean and variance equals to variance of the data. The reason to constraint realization to model parameters lies in linear nature of EnKF updating. If model parameters and state variables were uncorrelated, secondary data of state variables would not contribute to estimate of model parameters (porosity and permeability in our case). Next step is to add mean of predicted state variables to every SGS realization to get corresponding realization of state variable. Finally N_e realizations of state variables are obtained and entire reservoir model is defined. Note that even though number of flow simulation is reduced to one, geostatistical simulation should be run instead to create ensemble of realizations. Therefore, this concept should be applied only when computational cost of SGS is lower than cost of running flow simulation. Time is sacrificed against estimation quality, which unfortunately drops in comparison to results from conventional EnKF procedure. In proposed shortcut realizations of state variables are generated only by means of geostatistical tools in order to supply sufficient number of degrees of freedom, and, thus, should not be examined individually for estimation of state variable distribution. Only the mean of predicted state variables has a meaningful interpretation. This shortcut is an adaptation of geostatistical co-simulation for EnKF.

A comparison of conventional implementation of EnKF with proposed shortcut in algorithm is presented in a 2D SAGD example, where porosity field is estimated using porosity and temperature data. Base case of porosity is shown in the Figure 13, which comprises two distinct low and high values zones. Locations of porosity and temperature observations can be found in the Figure 14. It is clear that porosity data are sampled from single vertical surveillance well, while temperature gauges are allocated along two vertical wells and in well pair. Initial isotropic porosity ensemble and EnKF porosity best estimates obtained for three different cases are presented in the same figure. First estimate is derived using only high value porosity data, and, hence, low porosity zone is not captured at all. Next porosity estimate is obtained using porosity and temperature observations with shortcut in EnKF implementation algorithm. It is obvious that low value zone is highlighted in the estimate. However,

unrealistic high porosity values are present at the bottom left corner. Conventional EnKF approach with porosity and temperature data is used to get third estimate. It is argued that last estimate reproduces base case much better than any previous ones. In all cases the data are matched pretty well. Even though shortcut does not have the same estimation quality as conventional approach, it can be still applied for estimation of distribution of a static variable, when computational overburden is an issue. Discrepancy of EnKF estimates from the Figure 13 and Figure 14 is due to different semivariogram model selected for generation of initial porosity ensemble, what proves once again importance of proper choice of initial ensemble.

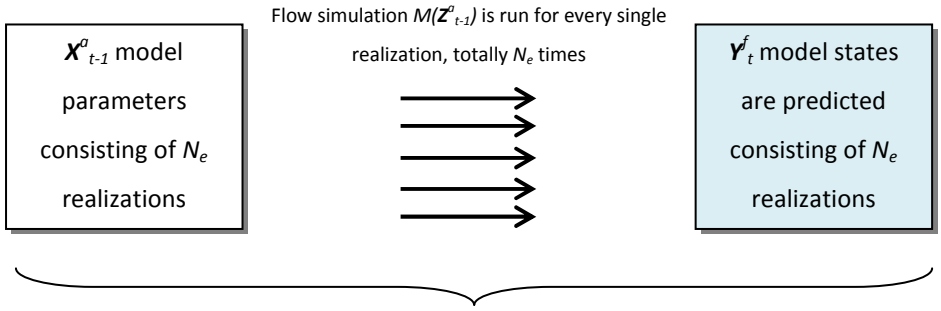


Figure 2: Conventional forecast step of EnKF from time step *t-1* to *t*

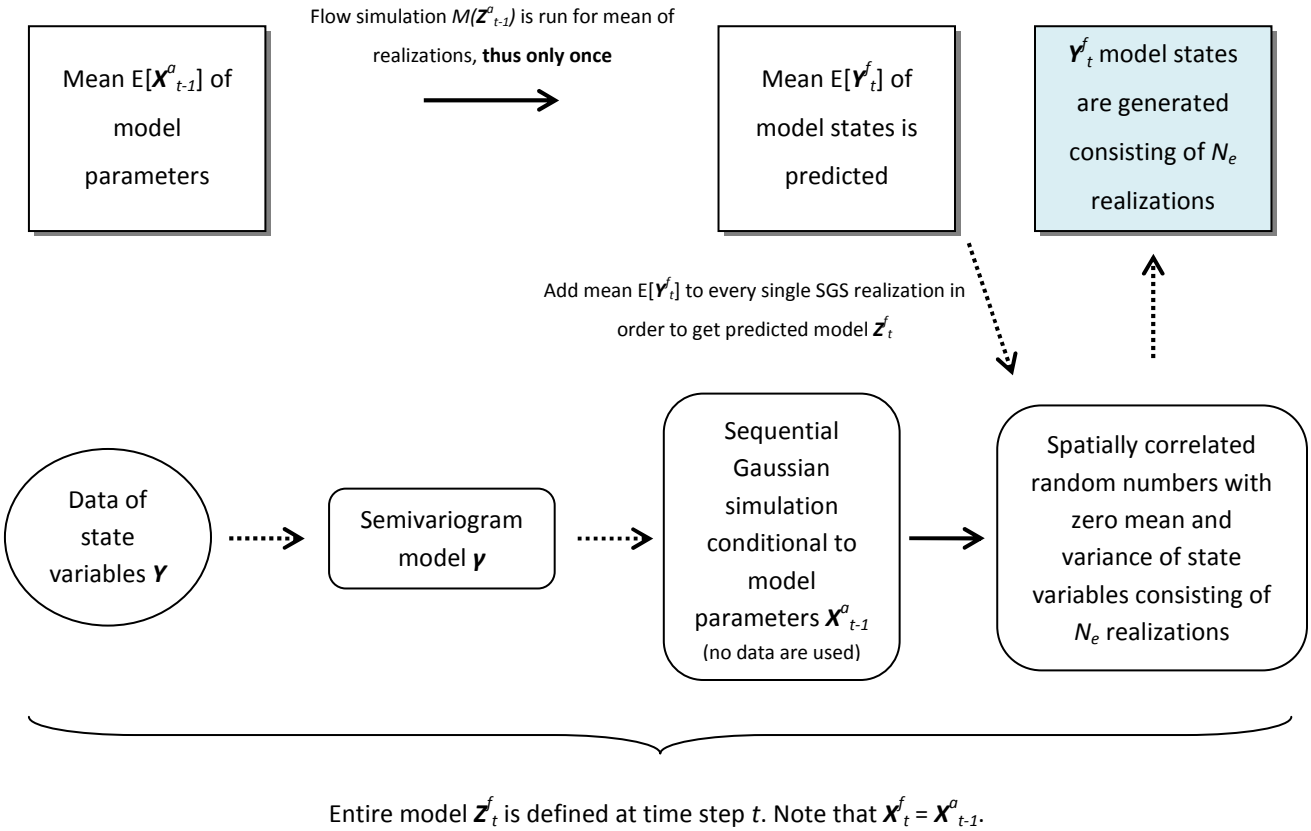


Figure 3: Proposed forecast step of EnKF based on mean of realizations concept from time step *t-1* to *t*

5. Program `enkf.exe`

In order to run program `enkf.exe` its parameter file should be defined first and several supplementary executable files have to be stored in folder `enkfsupexe`, which must be present in the program directory. Supplementary programs are `poroperm.exe` (see the Appendix for program description), `starstogslib.exe`, `fluidsub_tp.exe`, `seisdif.exe` (Zagayevskiy and Deutsch, 2011), and `sgsim.exe` (Deutsch and Journel, 1998). First program derives log permeability values from porosity. Second program converts STARS' output file to GSLib format. Third and fourth programs are devoted for generation of synthetic seismic attributes, which are a function of reservoir pressure and temperature. Last executable implements sequential Gaussian simulation. Difference in P-wave acoustic impedances ΔZ_p is chosen as a seismic attribute due to its ability to capture geological features and propagation of steam chamber growth (Zagayevskiy and Deutsch, 2011). Description of the parameter file of program `enkf.exe` is given below and sample parameter file is presented in the Figure 4.

```
                        Parameters for ENKF
                        *******************

       START OF PARAMETERS:
1-     1                                -EnKF: implementation option (1-conventional, 0-shortcut)
2-     100                              -  number of realizations
3-     50    0.5   1.0                  -  nx, xmin, xsize
4-     10   50.0 100.0                  -  ny, ymin, ysize
5-     30    0.5   1.0                  -  nz, zmin, zsize
6-     69069                            -  random number generation seed
7-     poro_initial.out                 -  input file with gridded initial porosity ensemble
8-     1                                -  column for attribute
9-     enkf.out                         -  output file name for EnKF estimates
10-    0.1                              -Observations: standard dev. of measurement error in %
11-    obs_poro.out                     -input file name with porosity data
12-    1   1                            -  available (1-yes,0-no), integration (1-global,0-local)
13-    1   2   3   4                    -  column for X, Y, Z coordinates and attribute
14-    3   3   3                        -  updating window nxup, nyup, nzup for local integration
15-    obs_perm.out                     -input file name with permeability data
16-    1   1                            -  available (1-yes,0-no), integration (1-global,0-local)
17-    1   2   3   4                    -  column for X, Y, Z coordinates and attribute
18-    3   3   3                        -  updating window nxup, nyup, nzup for local integration
19-    obs_temp.out                     -input file name with temperature data
20-    1   1   270.0                    -  available, integration (1-global,0-local), t in days
21-    1   2   3   4                    -  column for X, Y, Z coordinates and attribute
22-    5   3   3                        -  updating window nxup, nyup, nzup for local integration
23-    obs_difz.out                     -input file name with impedance difference data
24-    1   0   90.0    270.0            -  available, integration (1-global,0-local), t in days
25-    1   2   3   4                    -  column for X, Y, Z coordinates and attribute
26-    3   3   3                        -  updating window nxup, nyup, nzup for local integration
27-    1.0                              -Porosity-perm. model: correlation between Phi and log(K)
28-    2.5                              -  intercept of regression line, log(mD)
29-    2.5                              -  slope of regression line
30-    1.58   2.65                      -Petroelastic model: densities of mineral 1 & 2 in g/cm3
31-    0.7                              -  volume fraction of mineral 1
32-    15.0   37.0                      -  bulk moduli of mineral 1 and 2 in GPa
33-    6.0    44.0                      -  shear moduli of mineral 1 and 2 in GPa
34-    50000.0                          -  salinity of brine water in ppm
35-    0.8                              -  specific gas gravity (ratio of gas to air densities)
36-    0.9                              -  density of oil at standard (surface) condition
37-    25   1   3                       -SAGD: block numbers xloc, yloc, zloc of heel of producer
38-    25   1   3                       -  block numbers xloc, yloc, zloc of toe of producer
39-    25   1   7                       -  block numbers xloc, yloc, zloc of heel of injector
40-    25   1   7                       -  block numbers xloc, yloc, zloc of toe of injector
```

Figure 4: Parameter file of program `enkf.exe` for implementation of EnKF data assimilation technique

Parameter file is divided into five distinct sections: EnKF and model specifications section, section for observations, porosity-permeability sub model section, petroelastic sub model section, and section for SAGD production scheme. Thus, first section is devoted for EnKF and model grid. EnKF implementation option is defined on first line. It is either conventional approach or shortcut for reduction of number of times flow simulator is run at

forecast step as it is described in the previous section. Only ensembles of temperature and difference in acoustic impedances are computed using shortcut, where their initial values are generated unconditional to data, but correlated to initial ensemble of updated porosity values. Next line is reserved for number of realizations or ensemble size. Lines 3 to 5 determine model grid. Number of blocks, coordinates of origin and block sizes in three principal X, Y, Z directions are entered here (Deutsch and Journel, 1998). Random number generation seed is defined on line 6, which is used for multiple reasons, e.g. to generate permeability values based on porosity or to assign white noise to observed data at analysis step of EnKF. Input file with initial porosity ensemble is specified on next line 7, whose size has to be equal to number of realizations. It is recommended to use SGS for generation of initial ensemble with proper semivariogram model. Column for porosity attribute is defined on next line. Root name for output files is entered on line 9. The program generates several output files. The file names for EnKF estimates of porosity, permeability, temperature and difference in acoustic impedances have prefixes `poro_`, `perm_`, `temp_t1_`, and `difz_t0_t1_` respectively, where *t0* stands for time step when baseline survey is conducted and *t1* represents time step when temperature and repeated acoustic impedance are measured. Similar output files for best estimates and estimation variances of every variables are generated as well with intermediate prefix `_mv_` .

Second section is about data measurements. Standard deviation of data measurement error in percentage of true data value is defined on line 10. Next sixteen lines are used to define names of input files with data and type of assimilation procedure, which is either global or local, which is based on localization of updating matrix. Thus, name of input file with porosity observations is defined on line 11. Flags for porosity data availability and assimilation procedure are entered on the next line. Line 13 is for specification of columns with coordinates of porosity data and the attribute. If local assimilation method is chosen, half size of updating window for porosity data is entered on line 14 in units of blocks. Subsequent four lines are exactly the same as previous four, but only used for specification of permeability data. Definition of temperature and difference in acoustic impedances data is slightly different than for porosity and permeability data. Measurement time should be indicated for temperature and acoustic impedance, and furthermore, time step of baseline survey should be specified for acoustic impedance.

Third section of the parameter file is about porosity-log permeability relationship. Forecasted values of permeability variable are generated based on linear regression model (Equation (8)) and porosity values. Note that α and β from the Equation (7) are constant 1.0 and 0.5, respectively. Correlation coefficient between porosity and log permeability, intercept and slope of regression line are defined on lines 27 – 29.

Parameters of petroelastic model based on pressure and temperature dependent Gassmann's fluid substitution model are specified in fourth section of the parameter file. For more details on Gassmann's theory see (Zagayevskiy and Deutsch, 2011). Densities in g/cm^3 of constituent minerals of a rock are indicated on line 30. Volume fraction of any one mineral is entered on line 31. Volume fraction of other mineral is computed automatically as: volume fraction of second mineral = 1.0 – volume fraction of first mineral. Elastic bulk and shear moduli of both minerals in GPa are defined on next two lines. Lines 34, 35, and 36 are reserved for brine water salinity in ppm, specific gas gravity (gas to oil densities ratio, which is usually between 0.56 and 1.80), and density of oil (or bitumen) at standard conditions (15.6 ^{0}C, 0.1 MPa).

Last fifth section is used for specification of SAGD production scheme, and more specifically for single well pair location in space. Block numbers of producer's and injector's heels and toes in three principal X, Y, Z directions are defined on lines 37 – 40.

6. Case Study

Elaborated case study on integration of porosity and permeability core data and reservoir temperature and difference in seismic attributes in a 2D SAGD petroleum reservoir model using EnKF is presented in this section. The objective of the study is to demonstrate data assimilation through EnKF and to prove that adequate additional data improves reservoir quality. Even though porosity and permeability data from vertical surveillance wells, continuous temperature measurements and difference of P-wave acoustic impedances from baseline and subsequent surveys are used to get best estimate of all four variables and assess associated uncertainties, estimation of model parameters (porosity and permeability) is the main task of this case study. Objective function shown in the Equation (2) is used as a criterion for history match. Mean root squared error (*MRSE*) value is used to

assess estimation quality over entire model or, in other words, to validate the results. It is implemented in a program called `rsae.exe` (see the Appendix), from which absolute errors are computed as well.

$$RSE_{ireal} = \sqrt{\frac{\sum_{i=1}^{Nx}\sum_{j=1}^{Ny}\sum_{k=1}^{Nz}\left(z_{i,j,k,}^{true} - \overline{z}_{i,j,k,ireal}^{estimate}\right)^2}{N_x \cdot N_y \cdot N_z}} \tag{12}$$

$$MRSE = \frac{\sum_{ireal=1}^{Ne} RSE_{ireal}}{N_e} \tag{13}$$

where, RSE_{ireal} is the root squared error value computed for *ireal*th realization; *MRSE* is the mean root squared error or the average of RSE_{ireal} over realizations; N_X, N_Y, and N_Z are the number of blocks of a model in X, Y, and Z directions; N_e is the number of realizations or the ensemble size; $Z^{true}_{i,j,k}$ is the true or base case value of a variable in *i*, *j*, *k* block of a model; $\overline{z}^{estimate}_{i,j,k,ireal}$ is the *ireal*th estimated realization average of blocks adjacent in *i*, *j*, *k* block, which fall in specified smoothing window size, it is selected as 7 x 1 x 7 for this case study.

Program `enkf.exe` and other miscellaneous executable files are used for implementation of data assimilation technique based on EnKF. Two-dimensional model is selected because of two reasons: first, it is easier to explore the problem on smaller model size, and, second, most of fluid flow occurs in cross sections perpendicular to SAGD well pair. Model comprises only one SAGD well pair and case study can be easily extended to three dimensions and cover a single well pad or entire oil field. Model grid is 50 x 1 x 30 blocks of 1.0 m x 100.0 m x 1.0 m size. Schematic of the reservoir model is shown in the Figure 15. Injecting well is placed in block (25, 1, 7) above producing well, which is in block (25, 1, 3), with separation distance of 4.0 m. 2D model cuts the well pair, which is located in its middle. CMG's thermal flow simulator STARS is used to simulate reservoir temperature, pressure, and fluid saturations, which are used later on in petroelastic model based on Gassmann's fluid substitution model to generate synthetic seismic attributes (Zagayevskiy and Deutsch, 2011). Physical and elastic properties of constituent minerals and fluids are assumed to be known. Relationship between porosity and permeability is established using discussed porosity-log permeability model. It is assumed that these two variables are perfectly correlated and parameters of regression models are certain. Also SAGD production scheme and regime are known. The reservoir is preheated for 90 days in a region around SAGD horizontal wells before steam injection and production begin. The porosity and permeability data are sampled before any warm-up procedure from two vertical surveillance wells, whose number is 14 from each vertical well. 28 permanent temperature gauges are located along vertical surveillance bore holes as well as in SAGD well pair. The temperature is measured three times: just before production begins (time step 0 – 90.0 days), after 90 and 180 days of production (time steps 1 and 2 – 180.0 and 270.0 days). Geophysical baseline survey is conducted at time step 0 and subsequent two surveys are conducted at time steps 1 and 2, which match temperature measurement time. Reservoir coverage of seismic is even and extensive with 250 observation locations. Base cases of porosity, permeability, temperature and difference in P-wave acoustic impedances for time steps 1 and 2 are shown in the Figure 16. Associated observation locations are presented in the next Figure 17. Thus, 10 different cases are examined for data assimilation technique based on conventional EnKF algorithm with localization of updating matrix for seismic attributes only with window size of 7 x 1 x 7 blocks in order to find best estimates of porosity and permeability fields and to show that incorporation of additional data improves estimates. Initial normal scores are generated using isotropic semivariogram model shown in the Equation (14). Then normal scores are modified to original units of porosity with mean of 0.25 and variance of 0.14 to obtain initial porosity ensemble, which honors data distribution. The case studies are briefly described below:

1. no data are used in estimation, isotropic initial ensembles of porosity and permeability are best estimates in this case
2. only 28 porosity data from both surveillance wells are used
3. 28 porosity and 28 permeability data from both surveillance wells are used

4. 14 porosity and 14 permeability data from single (right) surveillance well are used, which do not capture low value zone of the base case
5. 14 porosity and 14 permeability data from single (right) surveillance well and 28 temperature measurements from time step 1 (180.0 days) are used
6. 28 temperature measurements from time step 2 (270.0 days) are integrated in an ensemble from previous case
7. 14 porosity and 14 permeability data from single (right) surveillance well and 250 difference in acoustic impedances from time step 1 (180.0 days) are used
8. 250 difference in acoustic impedances from time step 2 (270.0 days) is integrated in an ensemble from previous case
9. 14 porosity and 14 permeability data from single (right) surveillance well, 28 temperature measurements from time step 1 (180.0 days), and 250 difference in acoustic impedances from time step 1 (180.0 days) are used
10. 28 temperature measurements from time step 1 (270.0 days) and 250 difference in acoustic impedances from time step 1 (270.0 days) are integrated in an ensemble from previous case

$$\gamma(h) = 0.01 + 0.99 \cdot Exp_{\substack{a_1=60.0m \\ a_2=60.0m \\ a_3=60.0m}}(h) \tag{14}$$

Table 1: Comparison of porosity and permeability estimates derived using different data sets

Data	Total number of data	MRSE of porosity estimate	Rank* of visual goodness of porosity estimate	MRSE of permeability estimate	Rank* of visual goodness of permeability estimate
No data	0	0.192	10	1679.5	10
φ data from two surveillance wells	28	0.105	1,2	908.1	1,2
φ and K data from two surveillance wells	56	0.104	1,2	906.8	1,2
φ and K data from one surveillance well	28	0.179	9	1633.8	9
φ, K and T from one time step	56	0.145	8	1361.6	8
φ, K and T from two time steps	84	0.111	6	910.2	6
φ, K and ΔZ_p from one time step	278	0.151	5	1291.7	5
φ, K and ΔZ_p from two time steps	528	0.106	4	978.5	4
φ, K, T and ΔZ_p from one time step	306	0.131	7	1372.9	7
φ, K, T and ΔZ_p from two time steps	584	0.086	3	835.2	3

φ – porosity
K – permeability
T – temperature
ΔZ_p – difference in P-wave acoustic impedance
* – ranked #1 is the best estimate, ranked #10 is the worst estimate among available ten estimates

The comparison of estimation results are summarized in the Table 1 from visual and *MRSE* comparisons and shown graphically in the Figure 25 for *MRSE*. Best estimates of all variables are shown in the Figure 18 – Figure 24 as well. It is clear that incorporation of additional data improves estimation quality of both porosity and permeability fields, which are assimilated from different sources, from different time steps or both. It might be

argued that extensive seismic data brings more information to the model than temperature measurements. Even though large datasets of soft secondary data, like temperature and seismic, are available, additional hard data of porosity and permeability lead to more accurate estimates of pattern of spatial continuity of these two variables (compare estimates derived from using hard data from both surveillance wells to estimates from hard data from single well and all secondary sources). In all cases the soft data are matched pretty well for every time step due to nonlinear nature of examined system, while hard data are matched well even after all time steps. To sum up, flow zones and barriers are predicted well in this simple 2D example.

7. Conclusion

Successful assimilation of hard data (porosity and permeability) and continuous soft data (temperature and difference in acoustic impedances) in a petroleum reservoir model through EnKF in order to constraint distributions of porosity and permeability fields is presented in this paper on 2D synthetic SAGD case study. Any additional data, which is a function of model parameters, improved estimation of static porosity and permeability fields. Extensive seismic data bring more information to the model than temperature data from surveillance wells and SAGD well pair. It is argued and shown graphically that additional hard data may be a better source of information for understanding reservoir geology than soft data. Large number of realizations is used and localization of updating matrix is carried out in order to avoid ensemble collapse when extensive seismic data are assimilated. Other implementation details and shortcut to save computational time, which is based on combined effect of propagation of mean of state variables only at forecast step and introduction of associated variation through simulated $N(0,1)$ realizations conditional to model parameters, are discussed as well.

References

Batzle, M. and Wang, Z., 1992, Seismic properties of pore fluids, *Geophysics*, 64, 1396-1408

Butler, R., 1991, *Thermal Recovery of Oil and Bitumen*, Prentice Hall, Englewood Cliffs, New Jersey, 524 pp.

Chen, Z., 2007, *Reservoir Simulation: Mathematical Techniques in Oil Recovery*, SIAM, Philadelphia, 248 p.

Deutsch, C.V. and Journel, A.G., 1998, *GSLIB: Geostatistical Software Library and User's Guide*, Oxford University Press, New York, 2nd Ed., 369 pp.

Deutsch, C.V., 2002, *Geostatistical Reservoir Modeling*, Oxford University Press, New York, 372 pp.

Dong, Y., Gu, Y. and Oliver, D.S., 2006, Sequential Assimilation of 4D Seismic Data for Reservoir Description Using the Ensemble Kalman Filter, *Journal of Petroleum Science and Engineering* 53, 83 – 99

Duru, O.O. and Horne, R.N., 2009, Simultaneous Interpretation of Pressure, Temperature and Flowrate Data for Improved Model Identification and Reservoir Parameter Estimation, *2009 SPE Annual Technical Conference and Exhibition*, New Orleans, 1 – 19

Duru, O.O. and Horne, R.N., 2010, Joint Inversion of Temperature and Pressure Measurements for Estimation of Permeability and Porosity Fields, *SPE Annual Technical Conference and Exhibition*, Italy, 1 – 11

Eastwood, J., Lebel, P., Dilay, A. and Blakeslee, S., 1994, Seismic Monitoring of Steam-based Recovery of Bitumen, *The Leading Edge*, 242 – 251

Evensen, G., 2007, *Data Assimilation: The Ensemble Kalman Filter*, Springer Verlag, Berlin, 305 pp.

Fahimuddin, A., 2010, 4D Seismic History Matching Using the Ensemble Kalman Filter (EnKF): Possibilities and Challenges, *PhD Thesis*, University of Bergen, Norway, 116 pp.

Gassmann, F., 1951, Über die elastizität poröser medien: Vierteljahrsschrift der Naturforschenden Gesellschaft in Zurich, 96, 1-23. The English translation of this paper is available at http://sepwww.stanford.edu/sep/berryman/PS/gassmann.pdf

Gómez-Hernánez, J.J., Sahuquillo, A. and Capilla, J., 1997, Stochastic Simulation of Transmissivity Fields Conditional to Both Transmissivity and Piezometric Data – I. Theory, Journal of Hydrology 203, 162 – 174

Johnston, P.F., Andersen, G.R., Wachi, N., Lee, D.S., Martens, F.G. and Han, D.H., 1992, Integration of Seismic Monitoring and Reservoir Simulation Results for a Steamflood at South Casper Creek Oil Field, Wyoming, *67th Annual Technical Conference and Exhibition of the SPE*, Washington, DC, 513 – 519

Landrø, M., 2001, Discrimination between Pressure and Fluid Saturation Changes from Time-lapse Seismic Data, Geophysics, Vol. 66, No 3, 836 – 844

Li, Z. and Zhu, D., 2010, Predicting Flow Profile of Horizontal Well by Downhole Pressure and Distributed Temperature Data for Waterdrive Reservoir, *SPE 2009 Annual Technical Conference and Exhibition*, New Orleans, 296 – 304

Li, Z., Yin, J., Zhu, D. and Datta-Gupta, A., 2011, Using Downhole Temperature Measurements to Assist Reservoir Characterization and Optimization, *Journal of Petroleum Science and Engineering*

Lumley, D., 2001, Time-Lapse Seismic Reservoir Monitoring, *Geophysics*, Vol. 66, No 1, p. 50 – 53

Myrseth, I, 2007, Ensemble Kalman Filter Adjusted to Time-Lapse Seismic Data, *Petroleum Geostatistics Conference, Portugal,* European Association of Geoscientists and Engineers

Skjervheim, J.-A., evensen, G., Aanonsen, S.I. and Johansesn, T.A., 2005, Incorporating 4D Seismic Data in Reservoir Simulation Models Using Ensemble Kalman Filter, *2005 SPE Annual Technical Conference and Exhibition*, Texas, 1 – 11

Wu, Y-S., Lu, G., Zhang, K. and Bodvarsson, G.S., 2007, An Integrated Modeling Approach for Characterizing Multiphase Flow, Chemical Transport, and Heat Transfer in Fractured Reservoirs, *SPE Europec/EAGE Annual Conference and Exhibition*, London, 1 – 10

Yuh, S.H., 2003, *Time-Lapse Seismic Monitoring of Subsurface Fluid Flow*, PhD Dissertation, Texas A&M University, 126 pp.

Zagayevskiy, Y.V., Hosseini, A.H. and Deutsch, C.V., 2010a, *A Guidebook on Ensemble Kalman Filter for Geostatistical Applications*, Centre for Computational Geostatistics, Edmonton, 82 pp.

Zagayevskiy, Y.V., Hosseini, A.H. and Deutsch, C.V., 2010b, Characteristics of the EnKF for Geostatistical Problems, *Centre for Computational Geostatistics* 12, 125-1 – 125-20

Zagayevskiy, Y.V. and Deutsch, C.V., 2010, Constraining Geostatistical Realizations to Temperature Data with an EnKF, *Centre for Computational Geostatistics* 12, 205-1 – 205-12

Zagayevskiy, Y.V. and Deutsch, C.V., 2011, Temperature and Pressure Dependent Gassmann's Fluid Substitution Model for Generation of Synthetic Seismic Attributes, *Centre for Computational Geostatistics* 13, 210-1 – 210-19

Zhang, W., Youn, S. and Doan, Q., 2005, Understanding Reservoir Architectures and Steam Chamber Growth at Christina Lake, Alberta, by Using 4D Seismic and Croswell Seismic Imaging, *SPE/PS-CIM/CHOA International Thermal Operations and Heavy Oil Symposium*

Appendix – Description of programs `poroperm.exe`, `extrobs.exe`, `rsae.exe` and `standvar.exe`

Theoretical background and description of parameter files of programs `poroperm.exe`, `extrobs.exe`, `rsae.exe` and `standvar.exe` are presented in the Appendix. First program is written to generate log permeability values from porosity using linear regression probabilistic model. Second program is devoted for extraction of values at specific locations from a gridded data file. Third program is for computation of mean root square error or mean absolute error, which are used to assess quality of estimates by comparing to a base case. Last program either standardizes distribution of any variable or transforms distribution of values having zero mean and unit variance to a distribution with specified mean and variance.

In porosity-log permeability model it is assumed that porosity and log permeability with base 10 are correlated random variables, whose coefficient is known. Linear regression model for these two variables can be expressed as shown in the Equation (15), where ε is the white noise: random variable with zero mean and specified standard deviation, which has to be defined. This noise introduces probabilistic feature to deterministic regression model. Correlation between porosity and log permeability can be written as in the Equation (16). Since white noise is uncorrelated to rest of the variables, variance of log permeability and covariance of porosity and log permeability can be expressed through variances of porosity and noise and slope of the regression line *b* (see Equations (17) and (18)). Finally after trivial mathematical derivations variance of white noise is found to be a function of correlation coefficient and variance of porosity (Equation (19)). Thus, in order to derive permeability values from porosity intercept and slope of linear regression line and correlation coefficient between porosity and log permeability have to be specified. Variance of porosity is computed from the data. Generated permeability values are correlated to porosity to extent of defined correlation coefficient.

$$\log(K_{xx}) = a + b \cdot \varphi + \varepsilon \tag{15}$$

$$CORR\left[\log(K_{xx}), \varphi\right] = \frac{COV\left[\log(K_{xx}), \varphi\right]}{\sqrt{VAR\left[\log(K_{xx})\right] \cdot VAR\left[\varphi\right]}} \tag{16}$$

$$VAR[\log(K_{xx})] = b^2 \cdot VAR[\varphi] + VAR[\varepsilon] \tag{17}$$

$$COV[\log(K_{xx}), \varphi] = b \cdot VAR[\varphi] \tag{18}$$

$$VAR[\varepsilon] = b^2 \cdot \left(\frac{1}{CORR[\log(K_{xx}), \varphi]} - 1 \right) \cdot VAR[\varphi] \tag{19}$$

Sample parameter file of program `poroperm.exe` that generates permeability values from porosity is given in the Figure 5. Input file name with porosity values is specified on line 1. Next line is for column attribute. Correlation coefficient between porosity and lo permeability is defined on line 3. Intercept in log(mD) and slope of linear regression line are entered on lines 4 and 5 respectively. Random number generation seed is required to draw values of white noise, which is specified on line 6. Last seventh line is devoted for name of output file, which contains permeability, log permeability, and porosity values in three columns.

```
                        Parameters for POROPERM
                        ***********************

       START OF PARAMETERS:
1-     poro.dat                          -input file with porosity data
2-     1                                 -   column for attribute
3-     0.8                               -correlation coefficient between poro and log perm
4-     1.5                               -intercept of regression line, log(mD)
5-     2                                 -slope of regression line
6-     69069                             -random number generation seed
7-     poroperm.out                      -output file for porosity and permeability
```

Figure 5: Parameter file of program `poroperm.exe` for generation of permeability values from porosity

Sample parameter file of program `extrobs.exe` is shown in the Figure 6. In this paper the program is used to extract observations from gridded base case for further use in data assimilation based on EnKF. Input file name with gridded data is specified on first line. Column for attribute is entered on next line. Model grid specification is defined on lines 3 – 5. Name of file with X, Y, and Z coordinates of desired observation locations and associated column numbers are specified on lines 6 and 8. Number of first *N* observations to extract is indicated on line 7. If all coordinates are intended to be used in extraction of the data, -1 flag should be entered on this line. Output file name containing observations, their coordinates and standard deviations of measurement error in percentage of true values (line 10) is specified on line 9.

```
                        Parameters for ExtrObs
                        **********************

       START OF PARAMETERS:
1-     input.dat                         -input file with data
2-     1                                 -   column for attribute
3-     10  5.0  10.0                     -   nx, xmn, xsiz
4-     10  5.0  10.0                     -   ny, ymn, ysiz
5-     10  5.0  10.0                     -   nz, zmn, zsiz
6-     obs_loc.dat                       -input file with observation locations
7-     10                                -   number of observation locations, -1 for all
8-     1   2   3                         -   columns for X, Y and Z coordinates
9-     obs.out                           -output file for observations
10-    1                                 -   std. err. of observations in % of true value
```

Figure 6: Parameter file of program `extrobs.exe` for picking values at specific locations from a gridded data file

Parameter file of the program `rsae.exe` used to compute mean root square error (Equations (12) and (13)) or mean absolute error for 2D models is presented in the Figure 7. Absolute errors AE_{ireal} are computed similar to root squared errors as shown in the Equation (20). Name of input file with simulated data is defined on line 1 with specified column attribute on the next line. Number of realizations and model grid size are entered on

lines 3 and 4. Name of a file with gridded base case or reference case values should be indicated on line 5 with column attribute on the next line. Line 7 is reserved for size of smoothing window for averaging estimated values falling into window region. Names of output files for smoothed values, difference between smoothed estimates and true values and errors are specified on last three lines of the parameter file.

$$AE_{ireal} = \frac{\sum_{i=1}^{Nx}\sum_{j=1}^{Ny}\sum_{k=1}^{Nz}\left|z_{i,j,k,}^{true} - \overline{z}_{i,j,k,ireal}^{estimate}\right|}{N_x \cdot N_y \cdot N_z} \tag{20}$$

```
                    Parameters for RSAE
                    ********************

      START OF PARAMETERS:
1-    simulation.dat           -input file with input data
2-    1                        -   column number with attribute
3-    100                      -   number of realizations
4-    100   100                -   2D grid specification: nx, ny
5-    base_case.dat            -input file with reference/base case
6-    1                        -   column number with attribute
7-    5                        -size of square moving average window (odd number)
8-    average.out              -output file for smoothed values
9-    difference.out           -output file for difference between smoothed and base values
10-   rsae.out                 -output file for root squared and absolute differences
```

Figure 7: Parameter file of program `rsea.exe` for calculation of mean root square error and mean absolute error

Purpose of the program `standvar.exe` is to either standardize values of any random variable according to the Equation (21), that is mean and standard deviation of standardized values should be equal to zero and one respectively, or back standardize values with zero mean and unit variance to a distribution with target original mean and standard deviation as shown in the Equation (22). The sample parameter file of the program is presented in the Figure 8. Name of input file with data, which have to be altered, is specified on first line. Column attribute for the data is defined on next line. Flag of option for standardization or back standardization is found on line 3. If back standardization procedure is selected, target mean and standard deviation of final distribution have to be specified on lines 5 and 6. Otherwise they should be skipped. Finally, name of output file is entered on last line of the parameter file.

$$z^{stand} = \frac{z^{original} - m_z^{original}}{\sigma_z^{original}} \tag{21}$$

$$z^{original} = z^{stand} \cdot \sigma_z^{original} + m_z^{original} \tag{22}$$

where, z^{stand} is the standardized value of variable Z; $z^{original}$ is the original value of variable Z; $m_z^{original}$ and $\sigma_z^{original}$ are the average and standard deviation of original values of variable Z respectively.

```
                    Parameters for STANDVAR
                    ***********************

1-    START OF PARAMETERS:
2-    input.dat                  -input file with data
3-    1                          -   column for attribute
4-    1                          -standardize (1) or back standardize (2) values?
5-    1000                       -if (2): mean value
6-    100                        -if (2): standard deviation
7-    output.out                 -output file for old and new data
```

Figure 8: Parameter file of program `standvar.exe` for standardization or back standardization

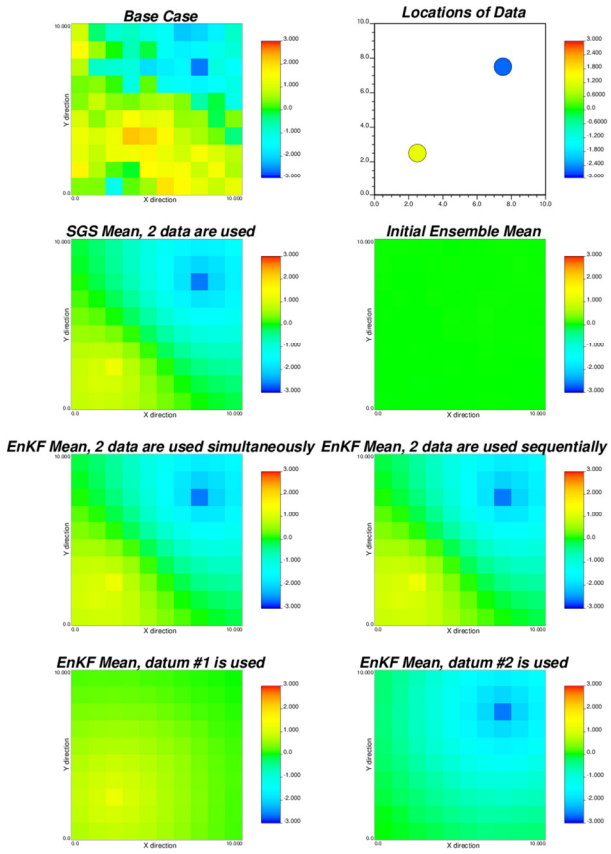


Figure 9: Comparison of best estimates of SGS and analysis step of EnKF, where data are assimilated simultaneously and sequentially

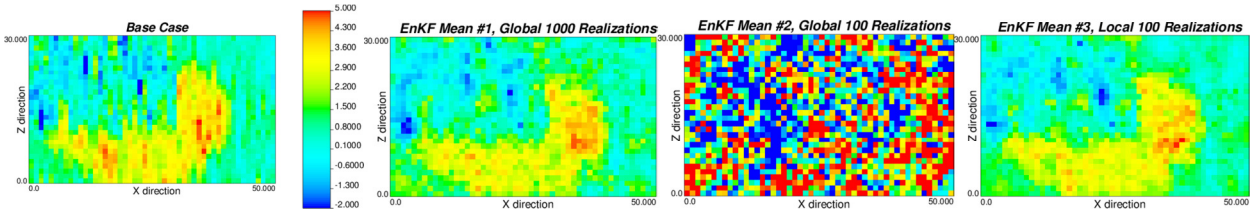


Figure 10: EnKF global and local estimates of seismic attribute: base case, global best estimate #1 from 1000 realizations, global best estimate #2 from 100 realizations, and local best estimate #3 from 100 realizations with localized updating matrix

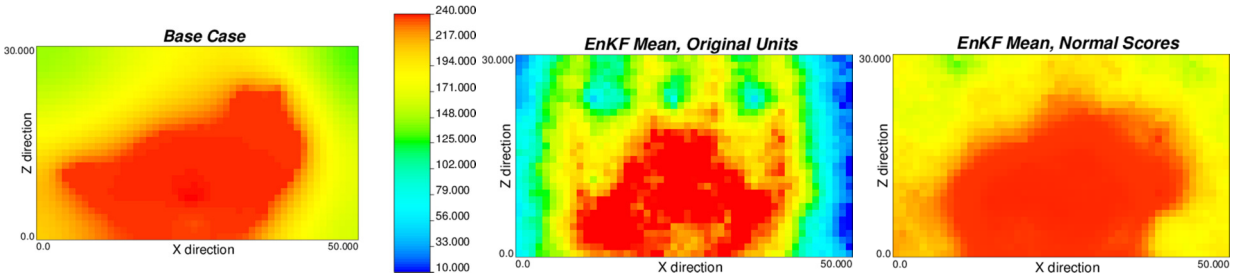


Figure 11: EnKF estimates of reservoir temperature: base case, mean of estimates at analysis step in original units and in normal scores

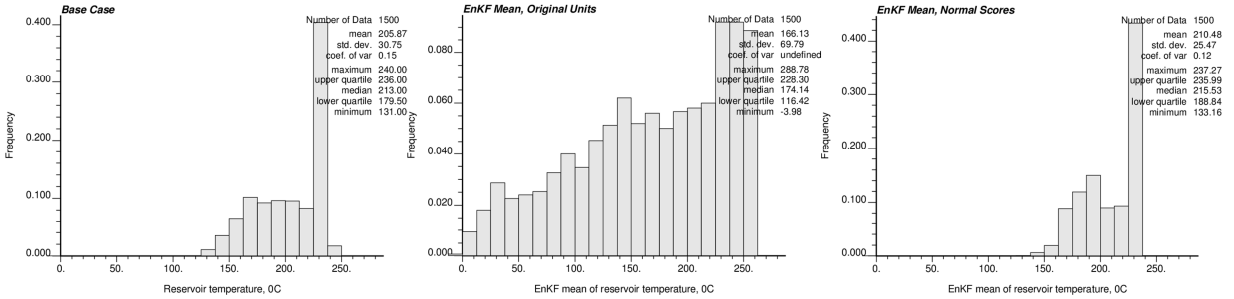


Figure 12: Histograms of EnKF estimates of reservoir temperature: base case, mean of estimates at analysis step in original units and in normal scores

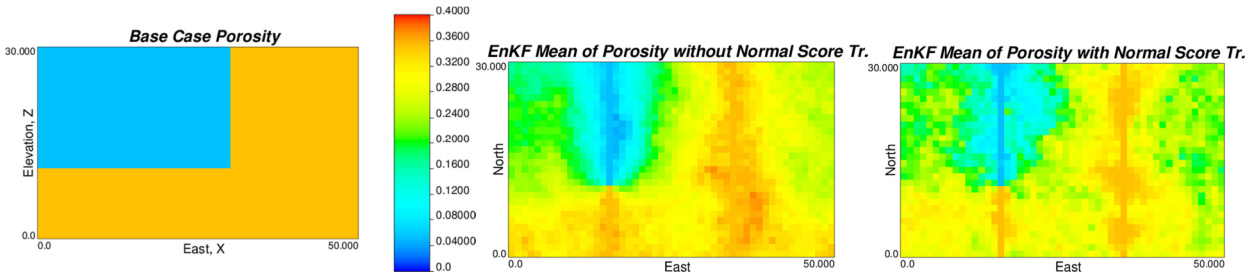


Figure 13: EnKF estimates of reservoir porosity: base case, mean of estimates at analysis step in original units and in normal scores

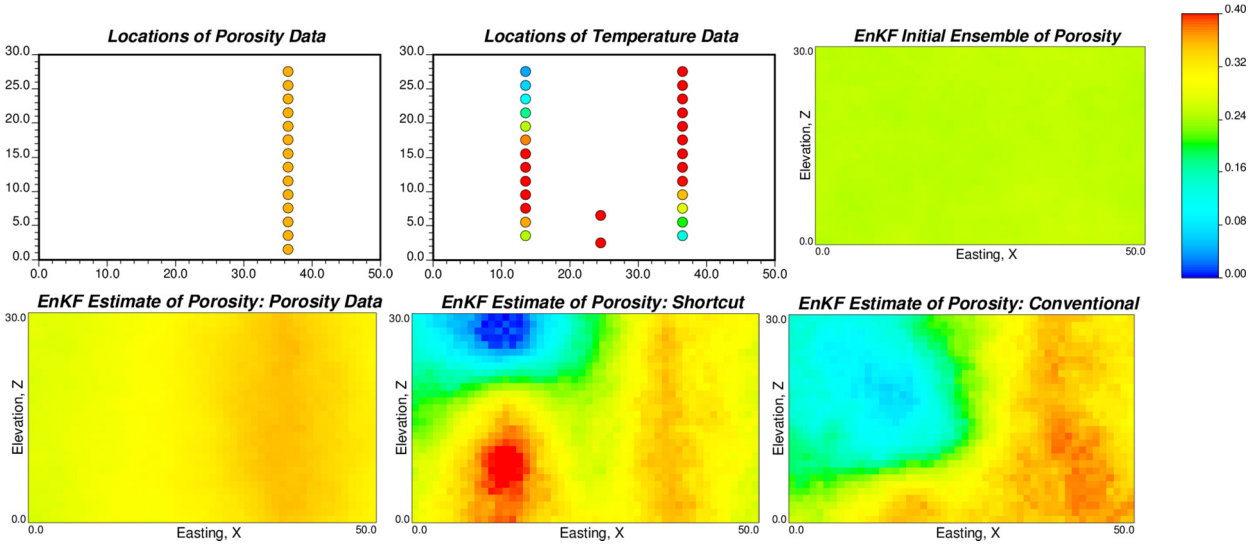


Figure 14: EnKF estimates of reservoir porosity: observation locations of porosity and temperature data, mean of initial porosity ensemble, mean of EnKF porosity estimate obtained from porosity data only, mean of EnKF porosity estimate obtained from porosity and temperature data using proposed shortcut, mean of EnKF porosity estimate obtained from porosity and temperature data using conventional filter algorithm. Base case is shown in the Figure 13.

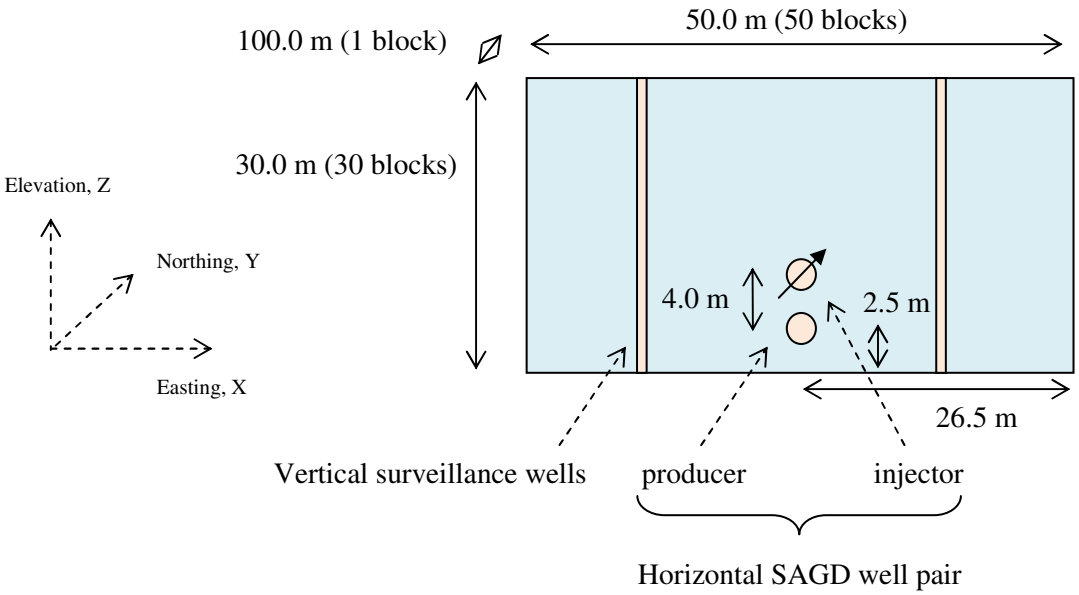


Figure 15: Schematic of 2D SAGD reservoir model

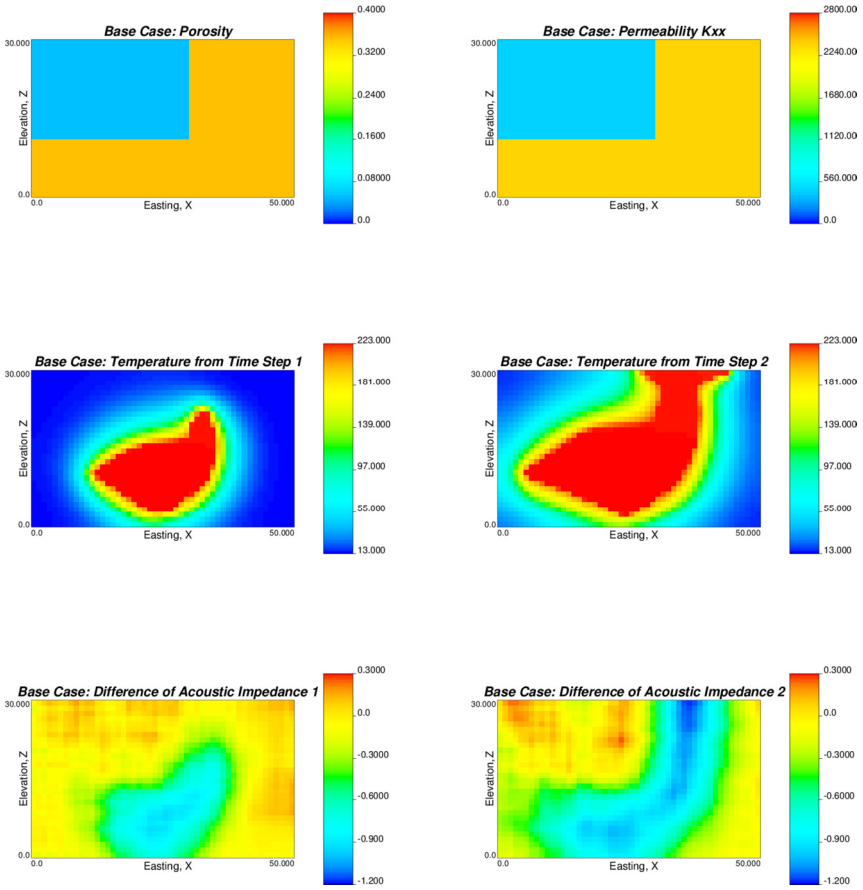


Figure 16: Base cases of porosity, tensor component of permeability in X direction, temperature from time steps 1 and 2, and difference of noisy P-wave acoustic impedances from baseline and subsequent surveys 1 and 2

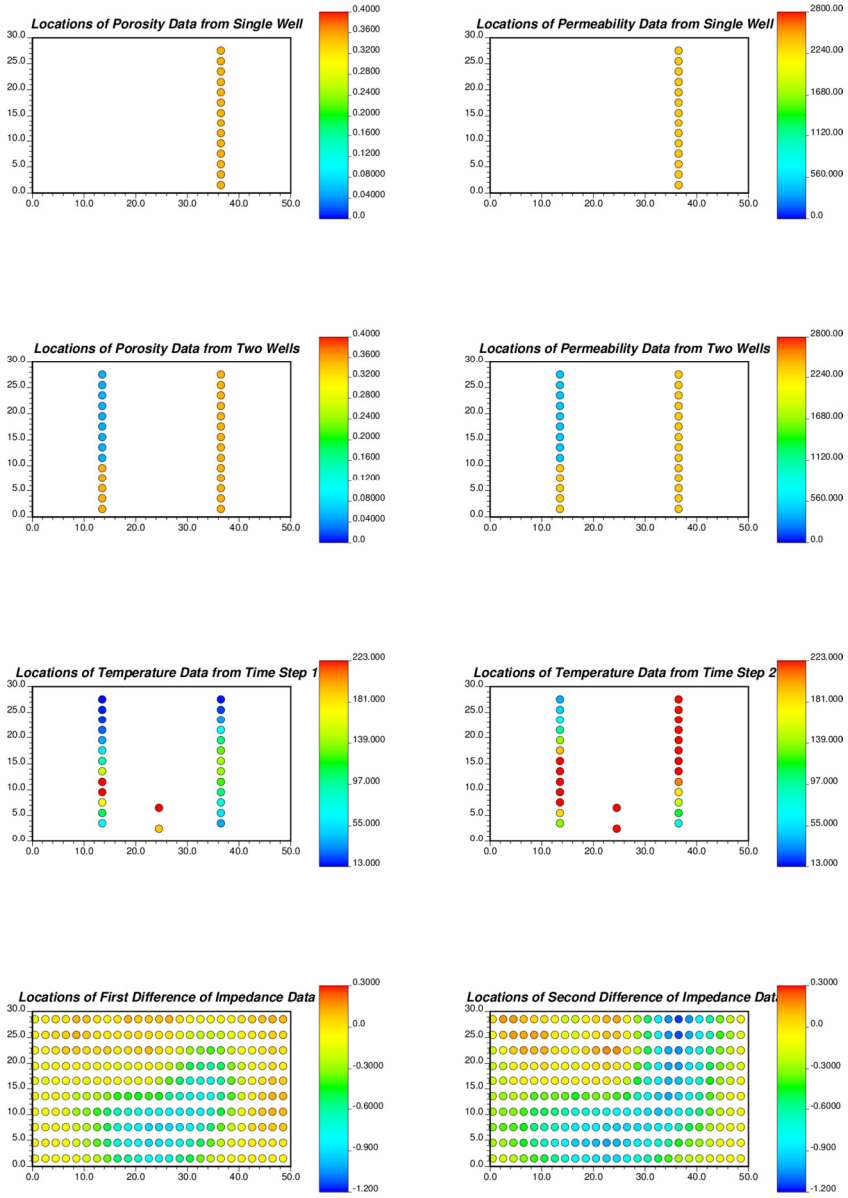


Figure 17: Observation locations of porosity and permeability data from one and two vertical surveillance wells, temperature data at two time steps from vertical wells and SAGD well pair, and difference of seismic attributes from three subsequent surveys sampled extensively over the field

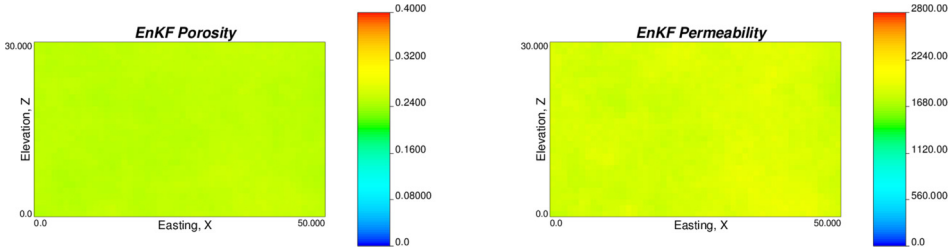


Figure 18: EnKF porosity and permeability best estimates using no data (initial ensembles)

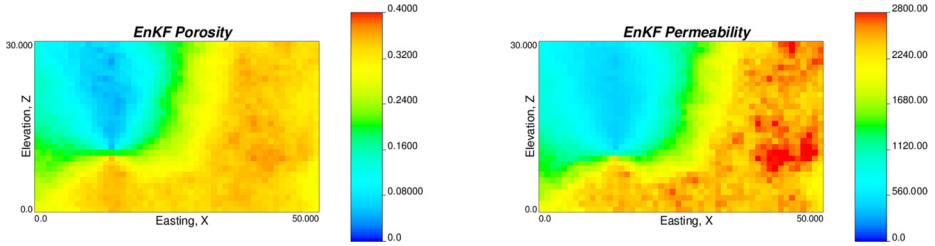


Figure 19: EnKF porosity and permeability best estimates using porosity data only from both surveillance wells

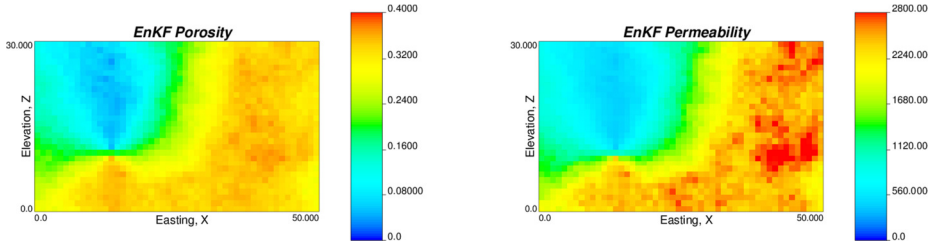


Figure 20: EnKF porosity and permeability best estimates using porosity and permeability data from both surveillance wells

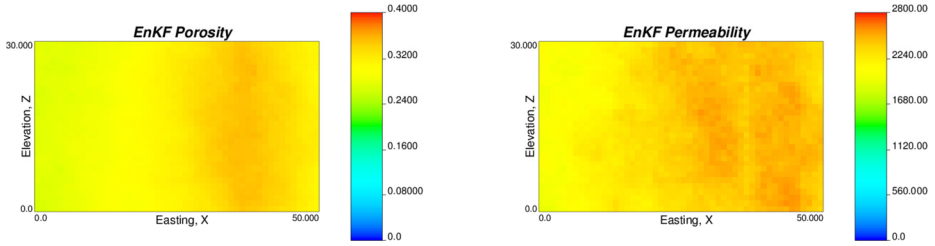


Figure 21: EnKF porosity and permeability best estimates using porosity and permeability data from single surveillance well

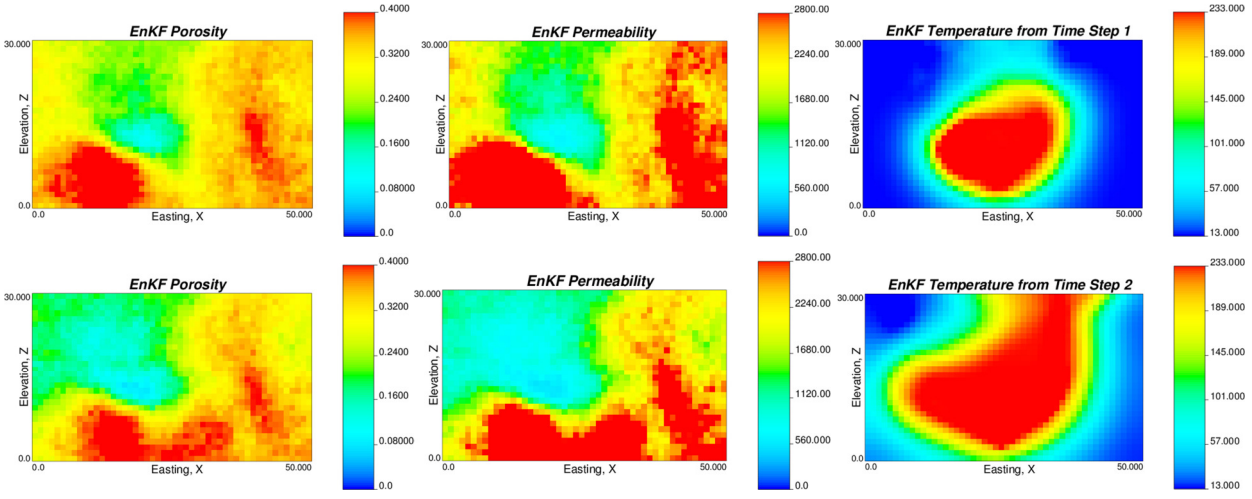


Figure 22: EnKF best estimates of porosity, permeability and temperature from two time steps using porosity, permeability data from single surveillance well and temperature measurements: top row figures are mean of EnKF estimates for porosity, permeability and temperature from step 1 data assimilation, bottom row figures are the same means, but when temperature from time step 2 is assimilated additionally

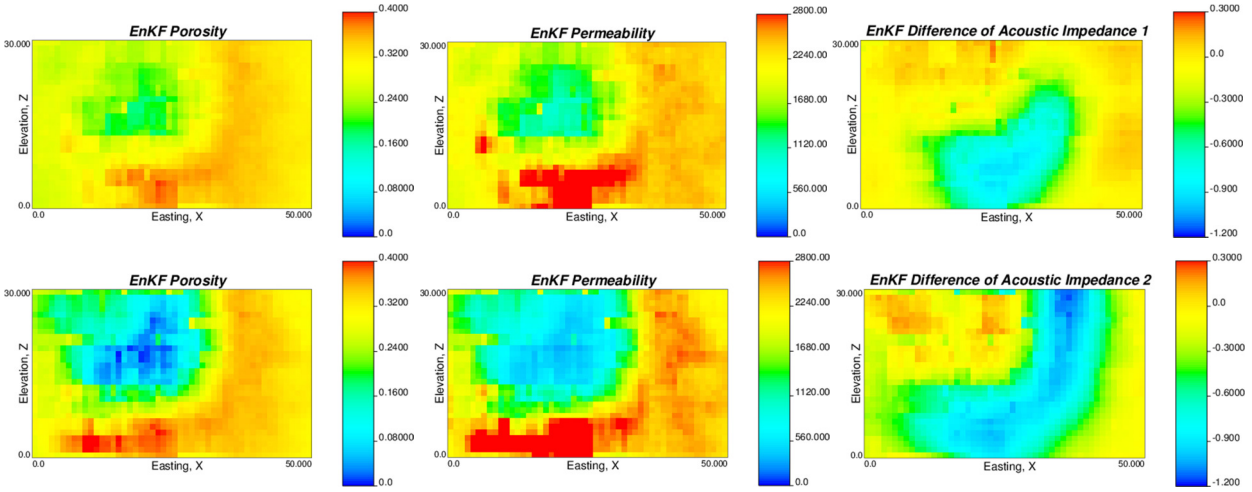


Figure 23: EnKF best estimates of porosity, permeability and difference in acoustic impedances from two time steps using porosity, permeability data from single surveillance well and difference in acoustic impedance from two surveys: top row figures are mean of EnKF estimates for porosity, permeability and acoustic impedance difference between time steps 0 and 1, bottom row figures are the same means, but when acoustic impedance difference data between time steps 0 and 2 are assimilated additionally

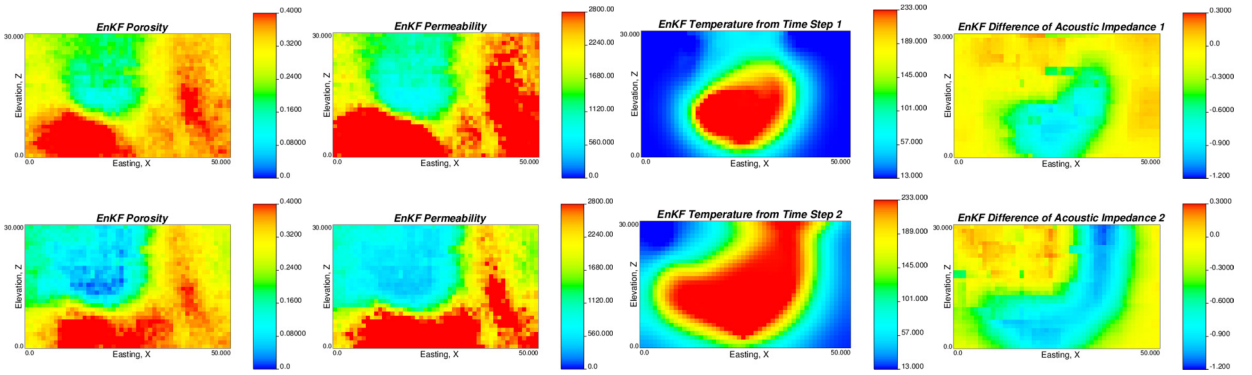
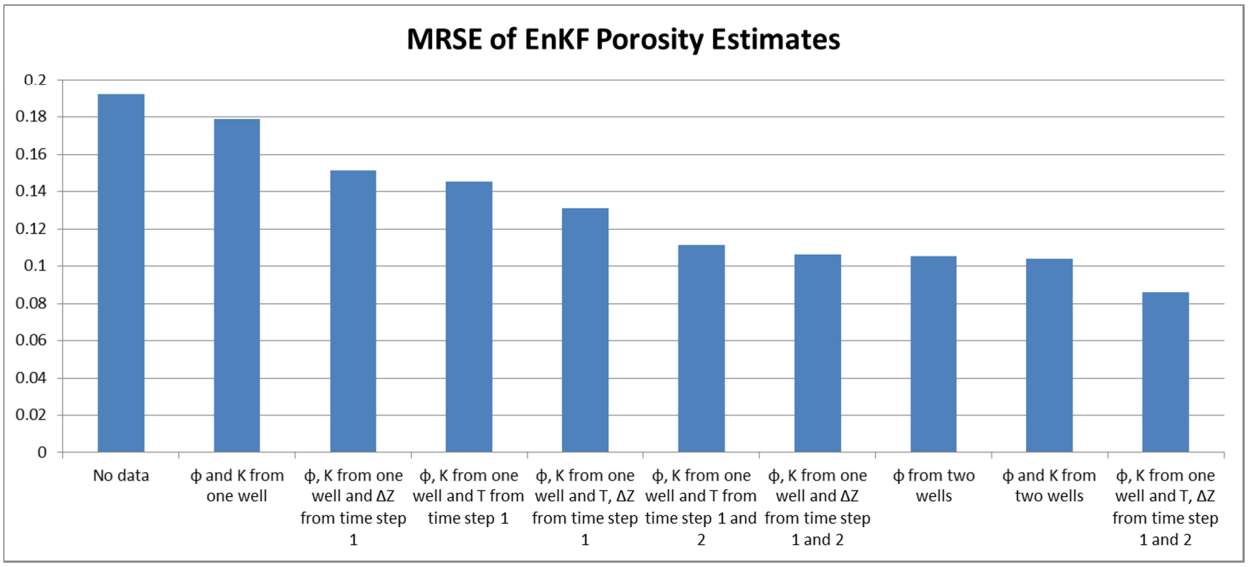


Figure 24: EnKF best estimates of porosity, permeability, temperature and difference in acoustic impedances from two time steps using porosity, permeability data from single surveillance well, temperature from both surveillance wells and difference in acoustic impedance from three surveys: top row figures are mean of EnKF estimates for porosity, permeability, temperature and acoustic impedance difference at time step 1, bottom row figures are the same means, but when temperature and acoustic impedance difference data are assimilated additionally at time step 2



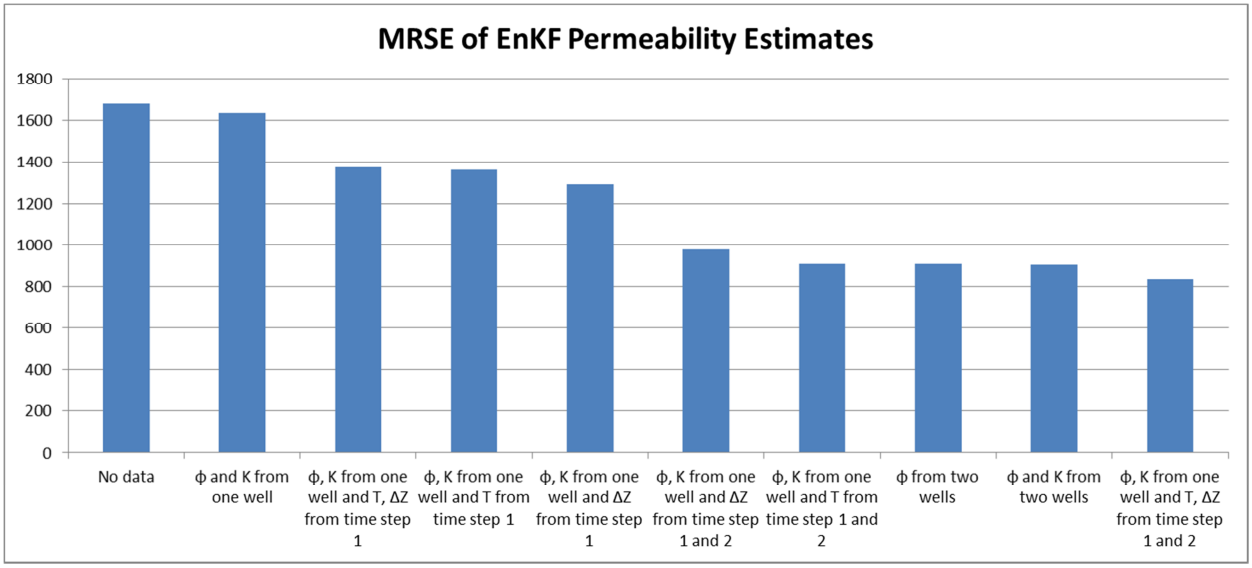


Figure 25: Comparison of *MRSE* of EnKF porosity (top chart) and permeability (bottom chart) estimates