

EnKF Applications to Thermal Petroleum Reservoir Characterization

Yevgeniy Zagayevskiy and Clayton V. Deutsch

This paper presents a summary of the application of the ensemble Kalman filter (EnKF) to petroleum reservoir characterization of thermally operated oil fields. The characteristics of the ensemble-based inverse modeling technique are assessed for northern Alberta bitumen reservoirs operated by steam assisted gravity drainage (SAGD) heavy oil extraction method. Proposed data assimilation technique effectively integrates static data such as petrophysical core data, and dynamic observations like continuous temperature observations or 4D seismic data into a petroleum reservoir model. Synthetic 2D and realistic 3D case studies demonstrate that additional information provides better insight into geologic properties of a reservoir and improves production forecasting. Theoretical background along with the case studies reveals strengths and weaknesses of the EnKF. The method performs well for linear or slightly nonlinear systems that follow nearly Gaussian distribution. The model updating procedure is fast, but forecasting may be computationally expensive especially for large systems. Localization techniques and replacement of the model realizations by their mean at forecast step are suggested to reduce computational time. Modifications of the method are required to successfully update categorical variables like facies. Even though the EnKF has been proven to be simple in implementation and effective for certain tasks, there is still room left for the research.

Introduction

The objective of a petroleum reservoir development project is profit maximization over the shortest possible period of time. Clear understanding of a petroleum reservoir geology helps to select appropriate technology for oil extraction, effectively place production/injection wells and surface facilities. All these and other activities lead to efficient field development. Learning dependence of production response on mutual relation between subsurface structure and properties of the extraction techniques will even further improve economics of the project. Therefore, there is a need to model petroleum system in conjunction with its static and dynamic components that can be presented in the form of petrophysical properties and time variations of physical conditions of a reservoir.

In order to plausibly describe behavior of the reservoir and predict future performance, various computational tools are available for petroleum reservoir characterization. Most of them use a concept of data assimilation, in which available data is integrated into the model by minimizing the difference between current model estimate and data at observation locations, while preserving overall physical and spatial structure of the model. Both static and dynamic data are integrated into the model to improve its quality (Oliver et al, 2008).

Reservoir characterization implies an inverse problem, where limited number of local data points is used to infer entire spatial configuration of the model (Oliver et al., 2008). A solution of the inverse problem is not unique, which means there is more than one solution that satisfies the problem settings by honoring data, and it is not obvious which solution is the closest to the reality (Tarantola, 2005).

All inverse modeling techniques can be divided into two broad groups based on strategy of minimizing mismatch between current model state and data: local techniques that search for solution around initial first-guess model estimate using Hessian and Jacobian matrices consisting of partial derivatives, and global techniques that search for solution globally trying to find the best set of parameters from available large pool of degrees of freedom. Local optimization techniques like Gauss-Newton method (Reynolds et al., 1996) and sequential self-calibration (SSC) technique (Gomez-Hernandez et al., 1997) require analytical description of the system, which is not an easy task. Global optimization techniques like genetic algorithm (Sen et al., 1995) and simulated annealing technique (Deutsch, 1992) might be more computationally expensive and might lead to wrong solution, if prior information is chosen poorly. However, they do not require prior construction of sensitivity matrices from analytical equations.

The ensemble Kalman filter or shortly EnKF is an inverse modeling technique that falls into global optimization pool, but it is based on set (ensemble) of equiprobable realizations of model estimates (Evensen, 2009). The method is based on Bayesian framework and Monte Carlo simulation (MCS). The

EnKF is easy to implement, but since it is based on ensemble of realizations, the technique is computationally expensive for complex large nonlinear systems. Its main advantage that attracted popularity to the technique is the ability to sequentially assimilate both static and dynamic data along with some prior information. Instead of deriving and solving analytical equations of petroleum reservoir behavior, existing numerical model can be adapted to predict reservoir performance. Its main drawback lies in the Gaussian nature of the estimate, i.e. distribution of the estimate tends to be normal. Moreover, it is hard to properly characterize highly nonlinear systems. Some modifications of the technique or transformation of the model distribution are necessary in order to implement the EnKF properly for categorical variables (Lorentzen et al, 2011).

This paper is organized as follows. First, family of Kalman filters is reviewed with description of strengths and weaknesses of each member. Second, theoretical background of the EnKF is presented. Third, characteristics and implementation details of the EnKF for petroleum reservoir characterization are shown. Fourth, methodology of EnKF application to thermally operated fields with available static and dynamic data is presented. Benefit of EnKF treatment as black box is explained. Fifth, synthetic 2D and realistic 3D case studies for SAGD fields are presented to show ability of the EnKF to assimilate static porosity and permeability data along with continuously measured temperature observations and time-lapse seismic attributes. Finally, conclusions are made and future work is defined.

Kalman Filter Family

History of EnKF origin is described in this section. Different modifications of Kalman filter (KF) inverse modeling technique are available in the literature. Originally, Kalman filter was devised by a Hungary-American mathematician and engineer Rudolf Kalman to remove noise from dynamic electrical signal (Kalman, 1960). In this context, the signal can be represented as a physical entity propagating in time and measured with error, which has to be removed in order to find the true value. Prior information is additionally used to further constrain the estimate of the truth by minimizing both measurement and prior estimate errors. Estimation is based on covariance structure of the system, and is a solution to Bayesian inversion problem.

The drawback of KF is that it can be applied only to linear systems. In reality, a few systems can be approximated in linear fashion. For this reason, the Kalman filter was modified to the extended Kalman filter (EKF), in which linear nature of the model is replaced with nonlinear form (Einicke and White, 1999). The problem with EKF is difficulty of deriving covariance matrix necessary for the estimation.

The ensemble Kalman filter (EnKF) is a stochastic extension of the EKF (Evensen, 2009). Model covariance matrix is replaced by sample covariance matrix, which is computed from the ensemble members also known as realizations. Each realization represents equiprobable estimate of the truth. The EnKF requires storing only a portion of the covariance matrix that describes relationship between model and data. This reduction of covariance matrix saves memory requirements and decreases computational time. However, a large ensemble is required to correctly approximate covariance structure, and propagation of the large nonlinear systems in time gets computationally expensive.

The unscented Kalman filter (UKF) is devised to reduce computational time of the forecast by propagating just several realizations in time and reconstructing rest of the realizations based on general knowledge of the system distribution form (Julier and Uhlmann, 1997). The reproduction of the distribution is not straightforward and complicated. Covariance structure is deviated from expected as ensemble evolves.

Minor modification of the EnKF, in which all available data from different sources and time are integrated at once, is called ensemble Kalman smoother (EnKS) (Evensen and van Leeuwen, 2000). It is a good alternative to the EnKF, if sequential propagation of the model is not required, for instance, when dynamic data is not assimilated frequently enough. Some other techniques similar to the EnKF are developed, but their application is limited in the petroleum literature. Randomized maximum likelihood (RML) and particle filter (PF) are few examples (Carrera and Neuman, 1986; Doucet et al., 2000). In the RML a large number of realizations are generated to represent all possible states of the modeled system. At data integration step, the realizations that honor the data the worst are discarded, other realizations are retained until deemed irrelevant, and final realization set is deemed to be a solution. Such approach is computationally expensive and does not guarantee that obtained solution is an optimal one.

Theoretical Background

In mathematical context, a petroleum reservoir is a complex system that can be described through a dynamic model \mathbf{z}_t . This model consists of two types of spatial functions that describe static and dynamic reservoir components: time insensitive model parameters \mathbf{m} and time t variant state variables \mathbf{s}_t . Both model parameters and state variables are stored in vector forms as shown in Eqs. (1) – (2.a). It is assumed that a state variable at current time step is a function of state variable value at previous time step, model parameters, initial and boundary conditions. Data vector \mathbf{d}_t is third supplementary component of the model.

$$\mathbf{z}(t) = \begin{bmatrix} \mathbf{m} \\ \mathbf{s}_t \end{bmatrix} \quad (1)$$

$$\mathbf{s}_t = F(\mathbf{s}_{t-1}, t \mid \mathbf{m}, \mathbf{s}_0, \mathbf{s}_t^b) + \boldsymbol{\varepsilon}_t \quad (2.a)$$

$$\mathbf{s}_0 = F(0 \mid \mathbf{m}) \quad (2.b)$$

$$\mathbf{s}_t^b = F(t \mid \mathbf{m}, A^b) \quad (2.c)$$

$$\mathbf{d}_t = G(t \mid \mathbf{m}) \quad (3)$$

where \mathbf{z}_t , \mathbf{m} , \mathbf{s}_t , and \mathbf{d}_t are the column vectors that represent entire model, model parameters, state variables, and data/observations respectively; t is the time index, and it is obvious that model parameters do not vary in time; \mathbf{s}_0 and \mathbf{s}_t^b are the initial and boundary conditions of the model state variables \mathbf{s}_t ; $\boldsymbol{\varepsilon}_t$ is the model error associated with the model propagation from time step $t - 1$ to next time step t ; A^b is the boundary domain; F and G are the model and observation operators that relate state variables and data to model parameter respectively. The operators may represent linear or nonlinear relationships depending on the system characteristics.

Generally, the petroleum reservoir system is numerically discretized in the space. Discrete values of continuous or categorical variables are assigned to the centers of the model grid blocks. Therefore, column vectors in Eqs. (1) – (3) compose of discrete values organized in some defined order.

The KF/EKF consists of two recursive steps: linear/nonlinear forecast step from one time step to another, and linear analysis step, which is based on a solution to Bayesian optimization problem. Analysis step is performed as many times as number of observations are assimilated from different time steps. Forecast step propagates state variables in time conditional to updated model parameters.

The forecast step can be derived as follows. If \mathbf{z}_{t-1} is viewed as some prior model estimate \mathbf{z}_t^a with zero model error, then model estimate \mathbf{z}_t at next time step conditional to updated model parameters \mathbf{m}^a becomes \mathbf{z}_t^f . The Eq. (2.a) also transforms to:

$$\mathbf{z}_t^f = F(\mathbf{z}_t^a \mid \mathbf{m}^a, \mathbf{s}_0, \mathbf{s}_t^b) \quad (4)$$

The Eq. (4) can be also called forward model. In most cases model error $\boldsymbol{\varepsilon}_t$ is ignored in Eq. (2.a), if the forward operator F is well defined. This equation is necessarily to establish plausible relationship between model parameters and state variables after data assimilation is performed.

The Bayesian framework for derivation of the analysis step is explained next. Bayesian rule can be expressed as shown in Eq. (5), where posterior probability distribution $p(\mathbf{z} \mid \mathbf{d})$ of the system is proportional to some prior probability $p(\mathbf{z})$ and data likelihood $p(\mathbf{d} \mid \mathbf{z})$ (Tarantola, 2005; Evensen, 2009). Time step is omitted, since updating step is performed without propagation in time. Model has distinct spatial structure before and after updating, which is updated based on spatial configuration and value of the integrated data. Essential probability density functions are presented in Eqs. (6) – (8) mathematically and assumed to follow normal (Gaussian) distribution.

$$p(\mathbf{z} | \mathbf{d}) \propto p(\mathbf{d} | \mathbf{z}) \cdot p(\mathbf{z}) \quad (5)$$

$$p(\mathbf{z} | \mathbf{d}) \propto \exp\left[-\frac{1}{2}(\mathbf{z} - \boldsymbol{\mu}_{z|d})^T \cdot (\mathbf{C}_{zz}^a)^{-1} \cdot (\mathbf{z} - \boldsymbol{\mu}_{z|d})\right] \quad (6)$$

$$p(\mathbf{d} | \mathbf{z}) \propto \exp\left[-\frac{1}{2}(\mathbf{z} - \boldsymbol{\mu}_{d|z})^T \cdot (\mathbf{C}_{dd})^{-1} \cdot (\mathbf{z} - \boldsymbol{\mu}_{d|z})\right] \quad (7)$$

$$p(\mathbf{z}) \propto \exp\left[-\frac{1}{2}(\mathbf{z} - \boldsymbol{\mu}_z)^T \cdot (\mathbf{C}_{zz}^f)^{-1} \cdot (\mathbf{z} - \boldsymbol{\mu}_z)\right] \quad (8)$$

where $\boldsymbol{\mu}_z$, $\boldsymbol{\mu}_{d|z}$, and $\boldsymbol{\mu}_{z|d}$ are the mean vectors of a prior, data, and posterior normal distributions of model \mathbf{z} ; and \mathbf{C}_{zz}^f , \mathbf{C}_{dd} , and \mathbf{C}_{zz}^a are the covariance matrices of a prior, data, and posterior normal distributions of model \mathbf{z} . Data covariance matrix \mathbf{C}_{dd} defines covariance structure of measurement errors.

When Eqs. (6) – (8) are used in Eq. (5) and maximum of posterior distribution is sought, equation for linear analysis step based on model covariance structure is derived. The output optimized posterior distribution defines updated model and covariance structure that honor newly assimilated data. The equations are shown below (Evensen, 2009):

$$\mathbf{z}^a = \mathbf{z}^f + \mathbf{K} \cdot (\mathbf{d} - \mathbf{H} \cdot \mathbf{z}^f) \quad (9)$$

$$\mathbf{C}_{zz}^a = (\mathbf{I} - \mathbf{K} \cdot \mathbf{H}) \cdot \mathbf{C}_{zz}^f \quad (10)$$

$$\mathbf{K} = \mathbf{C}_{zz}^f \cdot \mathbf{H}^T \cdot (\mathbf{H} \cdot \mathbf{C}_{zz}^f \cdot \mathbf{H}^T + \mathbf{C}_{dd})^{-1} \quad (11)$$

where \mathbf{K} is the Kalman gain or weighting matrix; \mathbf{I} is the identity matrix; and \mathbf{H} is the observation matrix that relates model estimates to data.

Same solution can be derived by looking at the problem at different angle. Objective function can be defined as a sum of weighted product of difference between estimate and truth for prior model and data. Here, prior model \mathbf{z}^f is viewed as a sum of model truth \mathbf{z}^{tr} and model errors $\boldsymbol{\varepsilon}_z$. Data \mathbf{d} is represented by sum of model truth \mathbf{z}^{tr} and measurement error $\boldsymbol{\varepsilon}_d$. Mathematically it can be expressed as in Eqs. (12) – (14). Minimization of this objective function implies bringing estimate \mathbf{z} closer to the truth \mathbf{z}^{tr} . Such minimization produces same Eqs. (9) – (11) (Evensen, 2009).

$$O = (\mathbf{z}^f - \mathbf{z})^T \cdot (\mathbf{C}_{zz}^f)^{-1} \cdot (\mathbf{z}^f - \mathbf{z}) + (\mathbf{d} - \mathbf{z})^T \cdot (\mathbf{C}_{dd})^{-1} \cdot (\mathbf{d} - \mathbf{z}) \quad (12)$$

$$\mathbf{z}^f = \mathbf{z}^{tr} + \boldsymbol{\varepsilon}_z \quad (13)$$

$$\mathbf{d} = \mathbf{z}^{tr} + \boldsymbol{\varepsilon}_d \quad (14)$$

where $\boldsymbol{\varepsilon}_z \sim N(0, \text{diag}(\mathbf{C}_{zz}^f))$ and $\boldsymbol{\varepsilon}_d \sim N(0, \text{diag}(\mathbf{C}_{dd}^f))$. It is assumed that model deviations $\boldsymbol{\varepsilon}_z$ and measurement errors $\boldsymbol{\varepsilon}_d$ are uncorrelated.

Similarities between KF's analysis equation and simple kriging are shown in (Zagayevskiy et al, 2010).

The EnKF differs from the KF/EKF in a sense that the model \mathbf{z} is represented by the matrix \mathbf{Z} , not vector, where all model realizations are stored. Covariance matrix \mathbf{C}_{zz}^f is carried forward through the ensemble members and, thus, covariance matrix \mathbf{C}_{zz}^f is approximated by sample covariance matrix $\hat{\mathbf{C}}_{zz}^f$:

$$\hat{\mathbf{C}}_{zz}^f \approx \mathbf{C}_{zz}^f = (\mathbf{Z}^f - \bar{\mathbf{Z}}^f) \cdot (\mathbf{Z}^f - \bar{\mathbf{Z}}^f)^T / (N_e - 1) \quad (15)$$

$$\bar{\mathbf{Z}}^f = (\mathbf{Z}^f \cdot \mathbf{e}^T) \cdot \mathbf{e} / N_e \quad (16)$$

where $\bar{\mathbf{Z}}^f$ is the model mean over ensemble members; \mathbf{e} is the $N_e \times 1$ row vector consisting of 1s; and N_e is the ensemble size, number of ensemble members or realizations.

Data vector \mathbf{d} is also extended into the matrix \mathbf{D} by adding some perturbation $\boldsymbol{\varepsilon}_d \sim \mathbf{N}(0, \mathbf{C}_{dd})$ to the existing data values. Perturbations are important, since true measurement error covariance matrix is not known, but has to be reproduced artificially to preserve validity of the EnKF equations.

Therefore, current model state and parameters are updated linearly at analysis step and are treated as input to the next forecast step. This process is continued until all observations are assimilated or objective function is optimized.

EnKF Characteristics and Implementation Details

Characteristics of the EnKF application have been studied and implementation details have been explored by the authors (Zagayevskiy et al., 2010a; Zagayevskiy et al., 2010b).

Beauty of the EnKF is that there is no need to fully understand the prediction mechanism at the forecast step. However, it is important that the tool used at the forecast step properly describes dynamic features of the studied system. Benefit of using full physics embedded into commercial flow simulator like CMG's STARS or Schlumberger's Eclipse versus proxy model is the same, in case if two simulators perform equally well. Therefore, black box in form of commercial software is not a bad idea and might be beneficial in terms of model accuracy, but computationally expensive in comparison with a proxy model.

The limitations and challenges are as follows. The EnKF is based on Gaussian framework, where updating step is linear and modeled variables are continuous. Thus, the technique is good for linear Gaussian systems. Simplification of the system to linear and transformation of variable distribution may be required in order efficiently implement EnKF. Updating of categorical variables can be hardly performed without either modification of the EnKF's analysis step or transformation of categorical variables to continuous form. Dovera and Rossa have introduced concept of Gaussian mixture models (GMM). Continuous properties of a model are grouped based on the affiliation to the model categories (Dovera and Rossa, 2007). Multimodal prior distribution of continuous variables is assumed, where each modal distribution behaves in a Gaussian manner and corresponds to a specific category. This approach has several drawbacks. There is a strong assumption of linear relationship between the model variables and data. Also the GMM technique is not capable of changing the shape and modal number of the prior distribution at the analysis step. Therefore, only continuous variables are updated, but categorical ones are left unchanged and inferred from updated continuous variables. The EnKF algorithm is modified to an extent when it is not simple anymore to implement. Another study on categorical variable updating with the EnKF was done by Lorentzen et al., in which categorical variables are represented by distance functions (Lorentzen et al., 2011). In this approach, prior information has stronger influence on the estimate than data. Therefore, if initial ensemble is chosen poorly, newly assimilated data hardly bring new information into the model. Application of the EnKF to categorical variables still remains an active research direction (Liu and Oliver, 2005).

Two main known implementation issues exist. First concern is related to ensemble collapse, when number of integrated data is larger than ensemble size, e.g. when seismic data is assimilated into a model. Second, computational time may be very high at the forecast step, if the modeled system is complex, and it is expensive to forecast even single realization. Model localization techniques are useful to solve these issues. Shortcut, in which entire ensemble propagation at the forecast step is replaced with ensemble mean/P50, saves computational time by sacrificing quality of the estimate.

If the ensemble size is not big enough, spurious covariance matrix occurs, which leads to ensemble collapse. Two localization techniques are studied: covariance matrix localization and updating matrix localization. The covariance matrix localization produces better estimates; but it is ambiguous to select appropriate smoothing function. Localization of updating matrix is straightforward; but leads to the artifacts in the estimates. Mathematically, two localization techniques can be expressed as follows:

$$\hat{\mathbf{C}}_{zz}^{f,lc} = \mathbf{L}^c \circ \hat{\mathbf{C}}_{zz}^f \quad (17)$$

$$\hat{\mathbf{C}}_{zz}^{f,lu} = \mathbf{L}^u \cdot \hat{\mathbf{C}}_{zz}^f \quad (18)$$

where \mathbf{L}^c and \mathbf{L}^u are the localization matrices for covariance and updating matrices localization procedures, respectively; \mathbf{L}^c has the same size as covariance matrix $\hat{\mathbf{C}}_{zz}^f$ and represents a smoothing function, values of which varies between 0 and 1: function takes value of 1 at data location and gradually decreases to 0 as estimate location is moving away from the sampled location; \circ is the element-wise or Hadamart product of matrix elements; \mathbf{L}^u also has the same size as $\hat{\mathbf{C}}_{zz}^f$, but it consists of either 0 or 1 values: 1 stands for the locations at which updating should take place, and 0 is used for the locations, values at which do not change with data assimilation.

The EnKF is very expensive at forecast step. A procedure based on mean of realizations is developed to decrease computational time at the forecast step. Ensemble mean of state variables is computed and used at the forecast step instead of the entire ensemble realizations. State variable realizations are reconstructed after forecast step by adding spatially correlated random fields to the forecasted mean. These random fields should be conditioned to model parameters and follow normal distribution with zero mean and spatial function of the state variables. Ranking of realizations and working with only several representative quantiles at the forecast step is another possibility to decrease computational time and make the technique more appealing in practice. Entire ensemble is reconstructed back after the forecast step is performed based on ranking, known shape of the distribution, and selected forecasted realizations.

Some other minor comment is to be careful in setting up EnKF algorithm. Even though EnKF equations are simple, it might be hard to set up a workflow, since various software should be linked together in order to relate all variable types. Convergence of the realizations to data values occurs as long as everything is set up correctly, there is no error in forward model, and measurement errors represented by a perturbation matrix is relatively small.

Model error $\boldsymbol{\varepsilon}_t$ is important; however, it is usually assumed that the forward model is well defined and the error is ignored. See more details in Evensen (Evensen, 2009).

EnKF Application to SAGD Petroleum Reservoir Characterization

Schematically data integration with the EnKF for SAGD petroleum reservoir characterization can be presented by a chart in Figure 1. The EnKF reservoir characterization technique consists of two main recursive steps: forecast step and analysis step. The technique can be viewed simply as state variable prediction at the forecast step and estimation of model parameters through data integration at the analysis step. It is important to estimate vertical permeability in SAGD fields, since it determines communication between an injector and producer, and vertical flow from reservoir to production wells.

Key model parameters are porosity, horizontal and vertical permeability. Main state variables are reservoir temperature change over time and difference between acoustic impedance from base line survey and subsequent monitoring surveys. The objective is to estimate model parameters by integrating core data, continuous temperature and exhaustive time-lapse seismic observations into petroleum reservoir model with the EnKF.

The algorithm starts with generation of initial porosity ensemble that consists of N_e realizations and contains some prior information about porosity distribution like spatial structure in form of semivariogram model or conceptual model of facies distribution. Simulation of porosity conditional to hard data is not required to generate initial ensemble, but such ensemble would be a good starting point.

Then, the initial porosity model is used at the forecast step, and realizations of all remaining key variables are simulated. Model operator at the forecast step comprises three sub models: porosity-permeability relationship, thermal flow simulation, and petroelastic model. Logarithmic values of horizontal permeability are generated according to porosity-permeability transform in Eq. (19) and vertical permeability is derived from Eq. (20). An example of porosity-logarithmic horizontal permeability function is shown in Figure 6. Realizations of reservoir pressure, temperature, and water, oil, and gas saturations at time step t are computed by thermal flow simulator, which is in this work presented by CMG's STARS. This thermal flow simulator is treated as a black box, parameters of which are assumed to be constant and known to a modeler. Since flow simulation is performed for each realization, the forecast step is computationally expensive for large systems. Realizations of acoustic impedance values for initial time step t_0 and current time step t are calculated through pressure- and temperature-dependent Gassmann's fluid substitution model using porosity and thermal flow simulator's output as input into this petroelastic model (Kumar, 2006; Zagayevskiy and Deutsch, 2011b). Difference in acoustic impedance is derived as subtraction of seismic attributes sampled at base case survey from monitoring survey seismic attributes. It is found that acoustic impedance is sensitive to porosity and temperature at comparable degree.

$$\log_{10}(K_h) = \log_{10}(K_{h,\min}) + a_1 \cdot \phi + a_2 \cdot [1 - \exp(-3 \cdot \phi / \phi_c)] \quad (19)$$

$$K_v = K_h \cdot R_{K_v/K_h} \quad (20)$$

where K_h and K_v are the horizontal and vertical permeability respectively, mD; ϕ is the porosity; $K_{h,\min}$ is the intercept of minimal hypothetical horizontal permeability value, when effective porosity is zero; ϕ_c is the critical porosity or porosity value when permeability stabilizes on log permeability-porosity curve; a_1 and a_2 are the experimental regression coefficients; and R_{K_v/K_h} is the experimental vertical permeability to horizontal permeability ratio.

At this point the entire model ensemble is defined and analysis step is performed next, if model parameter data or observations of continuous variables are available for current time step. The Eqs. (9) – (11) and (15) – (16) are used to linearly update model parameters and states through data assimilation and covariance function of the model. Then, the model is propagated in time to a next time step with updated model parameters. Thus, coupled forecast step and analysis step are run repeatedly until all data are assimilated or objective function is optimized.

Since studied SAGD reservoir system has a nonlinear nature, forecast step beside prediction ensures that forecasted state variables physically match updated model parameters. Some authors call this correction as a confirmation step. The issue is that after analysis step is performed, updated states are no longer correspond to updated model parameters and, thus, next run of the forecast step adjusts model states to updated model parameters.

Instead of running a flow simulation for each of N_e realizations at the forecast step, shortcut based on model parameter mean and unconditional simulation is proposed, which significantly decreases computational time with reduced quality of model estimate (Zagayevskiy et al, 2011a). In this shortcut, mean maps of porosity and permeability are computed and used in the flow simulation. It is assumed that resulting distribution of dynamic variables also represents mean of model states. Missing ensemble realizations of state variables are recreated through variability introduction to mean estimate of state variables by adding zero-mean unconditional realizations with spatial structure of model states correlated to primary model parameters. By doing so, physical appearance of the dynamic variables and their linear correlation with model parameters are preserved, which is important to compute Kalman gain weighting matrix. This approach is an implementation shortcut with less prediction accuracy and, thus, should be used when a full set of thermal flow simulation runs is computationally unaffordable.

Case studies shown in the next section highlight implementation details of the EnKF application to SAGD petroleum reservoir characterization, show contribution value of various data types, emphasize importance of initial ensemble, and address practical issues with the EnKF. The results show

improvements in reservoir performance prediction with additionally assimilated data. The FORTRAN program `enkf.exe` is available for reproduction of the results.

Case Studies

Two case studies are demonstrated to highlight properties and implementation details of the EnKF inverse modeling technique for estimation of static reservoir properties such as porosity and permeability and prediction of dynamic reservoir properties like cumulative oil production rate of SAGD operated fields. Different combinations of datasets are integrated into the model to assess importance and information value of each data type. Hard porosity and horizontal permeability data and soft temperature and seismic attributes are assimilated in a sequential fashion to improve prediction accuracy of the cumulative oil production rate.

Higher attention is paid to estimation of porosity and permeability properties of the reservoir, since their spatial configuration along with geological faults determine flow paths, baffles, and barriers for fluid flow from reservoir to production wells. Better understanding of model parameters, which are enhanced by the EnKF-based data assimilation, leads to more accurate prediction of dynamic reservoir performance.

The 2D conceptual model of channelized system is studied in first case study to examine ability of the EnKF to differentiate between flow paths and barriers. The 3D case study is shaped after conceptual model of north Alberta oil sands depositions in the estuarine environment to assess practical application of the EnKF.

For both case studies true distribution of model parameters and state variables is known in order to validate estimates and assess information value of different data types. Root mean square error (RMSE), mean standard deviation (MSD), and linear correlation coefficient (CC) are used to assess estimation accuracy of porosity field. Corresponding equations for porosity validation are shown in (21) – (23). The RMSE and CC indicate how close ensemble realizations are to the truth. The MSD shows how similar ensemble realizations are. Therefore, lower values of RMSE and MSD show better match of estimate to the truth and convergence respectively. Value of CC closer to 1.0 indicates better match between ensemble mean and the truth. Prediction accuracy of the estimated model is assessed through comparison of cumulative oil production of base case to the predictions from the EnKF. Since permeability field behaves identically to porosity field due to their well-defined transform in Eqs. (19) and (20), validation of permeability is deemed to be redundant.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N_b} \sum_{j=1}^{N_e} (m_{i,j} - m_i^{tr})^2}{N_e \cdot N_b}} \quad (21)$$

$$MSD = \sqrt{\frac{\sum_{i=1}^{N_b} \sum_{j=1}^{N_e} (m_{i,j} - \bar{m}_i)^2}{N_e \cdot N_b}} \quad (22)$$

$$CC = \frac{\sum_{i=1}^{N_b} (m_i - \bar{m}) \cdot (m_i^{tr} - \bar{m}^{tr})}{\sqrt{\sum_{i=1}^{N_b} (m_i - \bar{m})^2} \cdot \sqrt{\sum_{i=1}^{N_b} (m_i^{tr} - \bar{m}^{tr})^2}} \quad (23)$$

where N_b is the total number of grid blocks in the model; N_e is the ensemble size or number of realizations in the ensemble; i and j are the indices over block and realization number respectively; $m_{i,j}$ is the model parameter estimate at the i^{th} location of j^{th} realization, which is porosity in our case; m_i^{tr} is the true value of model parameter at i^{th} location; and \bar{m} , \bar{m}_i , and \bar{m}^{tr} are the average (mean) of estimates

over model grid and ensemble realizations, ensemble realizations, and average of truth over model grid, respectively.

Results are shown below and indicate that EnKF performed well in both cases with more accurate estimation outcome in 2D case study. Importance of careful selection of initial porosity ensemble is obviously seen in 3D case study.

2D Case Study

Base case constitutes background high quality reservoir rock and imbedded low quality mud barrier of channel-like shape as shown in Figure 2. The model can be viewed as a system of two lithological facies: good quality sand and poor quality shale, which is typical lithology in the McMurray formation, northern Alberta. Generated binary facies model is populated with single value porosity and permeability for each facies (Figure 2) according to the regression model relationship in Eq. (19) and (20) (Figure 6). This porosity-horizontal permeability transform is a fair practical approximation for facies in McMurray formation (Deutsch, 2010).

The model grid consists of 51 x 30 blocks, sides of which are 1.0 m each. A SAGD pair is placed symmetrically in the bottom middle of the model with separation distance of 5.0 m. Two vertical observation wells are drilled to sample porosity and permeability hard data. Thermocouples are located at the observation wells, producer, and injector to continuously measure reservoir temperature. Seismic survey is conducted over the area. Acoustic impedance is inverted from the seismic response and covers entire area of model grid with 3 x 3 m resolution. The continuous observations of temperature and acoustic impedance are sampled three times: just before the production begins, and 240 and 480 days after production start. Timeline of the case study is shown in Figure 5. Thus, the model is updated three times: with porosity and/or permeability data at $t_0 = 0$, with temperature and difference of acoustic impedance at $t_1 = 240$ days, and with temperature and difference of acoustic impedance at $t_2 = 480$ days with the EnKF. Base cases of porosity and difference of acoustic impedance for times steps t_1 and t_2 are presented in Figure 3 and Figure 4 respectively. It is obvious that continuous data reflect geological structure of the system. Updated model parameters are used to forecast cumulative oil production rate for period of 6000 days (~16.7 years). Different cases are examined to compare different modeling techniques, assess importance of different data types, and track improvements of the model over assimilation time. Full list of the cases is stored in Table 1. Fifty realizations is the default ensemble size with localized covariance matrix.

Table 1: 2D case study: different cases. The defaults are 50 realizations with localized covariance matrix

#	Method	Data and Comments
1	Kriging	Porosity
2	SGS	Porosity
3	EnKF	No data, initial ensemble
4	EnKF	Porosity
5	EnKF	Porosity, 1000 realizations
6	EnKF	Porosity, no localization
7	EnKF	Porosity, localization of updating matrix
8	EnKF	Porosity and permeability
9	EnKF	Porosity data and temperature observations from time step 1
10	EnKF	Porosity data and temperature observations from time step 1 with shortcut
11	EnKF	Porosity data and temperature observations from time steps 1 & 2
12	EnKF	Porosity data and difference in acoustic impedance observations from time step 1
13	EnKF	Porosity data and difference in acoustic impedance observations from time step 1 & 2
14	EnKF	Porosity data, temperature, difference in acoustic impedance observations from time 1
15	EnKF	Porosity data, temperature, difference in acoustic impedance observations from time 1 & 2

Figure 8 shows estimates derived from various techniques and conditional to different data sets. Numerically results are compared through RMSE, MSD, and CC, bar charts of which are shown in Figure

10. Change of one realization of porosity from initial to estimate constrained to porosity data, change of variance of porosity estimate, and dependence of permeability on porosity are presented in Figure 7. All model estimates are used to forecast the cumulative oil production, which are also compared to the base case (Figure 9).

It is observed that additional integration of the data improves porosity estimate and prediction accuracy of cumulative oil production rate. Integration of soft dynamic data definitely enhances model quality. EnKF performs almost as well as kriging and SGS for hard data integration, but EnKF excels conventional mapping tools, when additional dynamic data is integrated. Seismic data has the largest contribution to the estimation quality, because of its larger spatial coverage. Ensemble collapse is observed, when ensemble size is small in comparison with the number of assimilated data and no localization technique is applied. If ensemble size increases, model estimate improves as well. Localization techniques improve model without increase of ensemble size. Covariance localization of the form in Eq. (24) improves model better than localization of updating matrix. Localization of updating matrix leads artifacts in the porosity estimation due to boundary effect of updating window. Proposed shortcut at the forecast step is applied successfully to integrate temperature data that leads to reduction in computational time and slightly decrease in the estimation quality. When redundant data is used in the estimation like permeability with porosity or temperature with acoustic impedance, prediction accuracy improves, but porosity estimate gets worse. Even though the ensemble convergence to single realization with increasing number of assimilated data is a good sign, ensemble collapse should be avoided.

$$L_i^c = \exp\left(-5 \cdot l_i / \sqrt{l_x^2 + l_y^2 + l_z^2}\right) \quad (24)$$

where L_i^c is the element of covariance localization matrix L^c ; l_i is the distance between data location and estimate in the space; and l_x , l_y , and l_z are the model dimensions in X, Y, and Z directions.

3D Case Study

The 3D case study is set up similar to the 2D case study. But more complex and geologically realistic base case is selected that describes tidally influenced fluvial depositional environment typical for the McMurray formation in northern Alberta (Hassanpour and Deutsch, 2010). The model grid is represented by 16 x 25 x 15 blocks of 50.0 m, 8.0 m, and 2.0 m size respectively, layout of which is shown in Figure 11 and Figure 13. Grid-free object-based facies model is generated and rasterized with different resolutions: high resolution model is used as an ultimate truth, and low resolution model represents secondary base case, at scale of which EnKF estimation and prediction are performed to reduce computational time. The resolution of rasterization of the facies model has an effect on reservoir performance prediction, but slight effect is observed on facies proportions and other geological properties of the system (Figure 12 and Figure 20). The facies model is populated with porosity and permeability values that are correlated to each other as shown in Figure 14. Histogram of base case porosity is shown in the same figure.

Two SAGD well pairs are drilled for the bitumen extraction, and three vertical observation wells are available for hard data sampling and continuous temperature response measurements. Temperature change is observed in well 1 and slightly in well 2, but not in well 3, which means that flow properties around wells 2 and 3 are not as good as around well 1. The EnKF-based algorithm is used to estimate porosity and permeability and forecast cumulative oil production for 6000 days. Two different initial porosity ensembles are used to emphasize importance of proper generation of the initial ensemble. First initial porosity ensemble follows Gaussian distribution with mean and variance of porosity data. Second initial porosity ensemble has a binary facies pattern identical to the base case, but different values of porosity and permeability are used in model population. Various combinations of data are looked at to estimate porosity field in order to detect the effect of data type on predictions. The examined cases are tabulated in Table 2.

Estimates derived from first initial ensemble for various cases are shown in Figure 15, and estimates obtained from second initial ensemble are presented in Figure 16. Histograms of the estimates can be found in next two set of plots: Figure 17 and Figure 19. Computed values of validation measures are visually presented in Figure 18. Figure 20 shows a forecast made by the EnKF for various data

combinations. When cumulative oil production rates and porosity estimates are compared for cases generated with different initial porosity ensembles, it is observed that second initial ensemble closer to the truth leads to better forecast. In both cases updated porosity field tends to be normal. Original binary-like porosity model becomes normal due to Gaussian nature of the EnKF background. This observation indicates a drawback of the EnKF for modeling multimodal systems. Additional temperature data does not improve model, since temperature from first time step is deemed to be informative enough. Proposed shortcut works worse in 3D case study, and leads to extreme values in porosity estimate, but predicts better dynamic states. Improvement oil production forecast is small, if initial ensemble is chosen close to the reality, and improvement is significant, if initial ensemble is chosen far from the reality (Figure 20).

Table 2: 3D case study: different cases. The defaults are 50 realizations with localized covariance matrix

#	EnKF Method	Data
1	Conventional	No
2	Conventional	Porosity
3	Conventional	Porosity, temperature 1
4	Conventional	Porosity, temperature 1 and 2
5	Shortcut	Porosity, temperature 1
6	Shortcut	Porosity, temperature 1 and 2

Conclusions and Future Work

It has been shown that the EnKF is very powerful inverse modeling technique, especially when modifications are applied to the technique like covariance localization or shortcut at the forecast step. Two case studies for SAGD applications have been demonstrated, in which EnKF was able to closely predict future performance of the reservoir and describe subsurface geological properties of the McMurray formation in northern Alberta. Additionally assimilated data improve quality of static properties of the model and prediction accuracy of cumulative oil production rate.

Commercial benefit of the EnKF applications to the SAGD fields should be studied further. Real case studies implementations are limited due to computational expense and inability to properly characterize categorical variables. Parallel programming, modifications, and simplifications should be considered for the future applications to reduce computational time. Developments towards modeling of non-Gaussian systems should be enhanced as well.

References

- Carrera, J. and Neuman, S.P., 1986. Estimation of Aquifer Parameters Under Transient and Steady State Conditions: 1. Maximum Likelihood Method Incorporating Prior Information, *Water Resources Research*, 22(2), 199 – 210.
- Deutsch, C.V. 1992. *Annealing Techniques Applied to Reservoir Modeling and The Integration of Geological and Engineering (Well Test) Data*, PhD Thesis, Stanford University, 325 p.
- Deutsch, C.V., Journel, A.G. 1998. *GSLIB: Geostatistical Software Library and User's Guide*: 2nd Edition. New York: Oxford University Press, 369 p.
- Deutsch, C.V. 2010. Estimation of Vertical Permeability in the McMurray Formation. *Journal of Canadian Petroleum Technology*, 49(12), 1 – 8.
- Doucet, A., de Freitas, N. and Gordon, N., 2000. *Sequential Monte Carlo Methods in Practice*, Springer, 591 p.
- Dovera, L. and Rossa, E.D. 2007. Ensemble Kalman Filter for Gaussian Mixture Models. *EAGE Conference*, Portugal, A16.
- Einicke, G.A. and White, L.B. 1999. Robust Extended Kalman Filtering, *IEEE Transactions on Signal Processing*, 47(9), 2596 – 2599.

- Evensen, G. and van Leeuwen, P.J. , 2000. An Ensemble Kalman Smoother for Nonlinear Dynamics, *Monthly Weather Review*, 128(6), 1852 – 1867.
- Evensen, G. 2009. *Data Assimilation: The Ensemble Kalman Filter: 2nd Edition*. Springer, 307 p.
- Gomez-Hernandez, J.J., Sahuquillo, A. and Capilla, J.E. 1997. Stochastic Simulation of Transmissivity Fields Conditional to Both Transmissivity and Piezometric Data – I. Theory, *Journal of Hydrology*, 203, 162 – 174.
- Hassanpour, R.M. and Deutsch, C.V. 2010. Grid-Free Object-based Modeling of IHS Facies in the McMurray Formation, *Centre for Computational Geostatistics: Report 12*, 203-1 – 203-5.
- Julier, S.J. and Uhlmann, J.K., 1997. *A New Extension of the Kalman Filter to Nonlinear Systems*, The Robotics Research Group, Department of Engineering Science, The University of Oxford, 12 p.
- Kalman, R.E. 1960. A New Approach to Linear Filtering and Prediction Problems, *Journal of Basic Engineering*, 82(1), 35 – 45.
- Kumar, D. 2006. A Tutorial on Gassmann Fluid Substitution: Formulation, Algorithm and Matlab Code, *Geohorizons*, 4 – 12.
- Liu, N. and Oliver, D.S., 2005. Ensemble Kalman Filter for Automatic History Matching of Geologic Facies, *Journal of Petroleum Science and Engineering*, 47, 147 – 161.
- Lorentzen, R.J., Nævdal, G. and Shafieirad, A. 2011. Estimating Facies Field using the Ensemble Kalman Filter and Distance Functions – Applied to Shallow-Marine Environments. *SPE Annual Conference and Exhibition*, Austria.
- Oliver, D.S., Reynolds, A.C. and Liu, N. 2008. *Inverse Theory for Petroleum Reservoir Characterization and History Matching*, Cambridge University Press, 380 p.
- Reynolds, A.C., He, N., Chu, L. and Oliver, D.S. 1996. Reparameterization Techniques for Generating Reservoir Descriptions Conditioned to Variograms and Well-Test Pressure Data, *SPE Journal*, 1(4), 413-426.
- Sen, M.K. Datta-Gupta, A., Stoffa, P.L., Lake, L.W. and Pope, G.A. 1995. Stochastic Reservoir Modeling Using Simulated Annealing and Genetic Algorithms, *SPE Formation Evaluation*, 10(1), 49-56.
- Tarantola, A. 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, 358 p.
- Zagayevskiy, Y., Hosseini, A.H. and Deutsch, C.V. 2010a. Characteristics of the EnKF for Geostatistical problems. *Centre for Computational Geostatistics: Report 12*, 125-1 – 125-20.
- Zagayevskiy, Y. and Deutsch, C.V. 2010a. Constraining Geostatistical Realizations to Temperature Data with an EnKF. *Centre for Computational Geostatistics: Report 12*, 205-1 – 205-12.
- Zagayevskiy, Y. and Deutsch, C.V. 2010b. Thoughts on Practical Inverse Modeling in Petroleum Reservoir Characterization. *Centre for Computational Geostatistics: Report 12*, 205-1 – 205-12.
- Zagayevskiy, Y., Hosseini, A.H. and Deutsch, C.V. 2010b. *Ensemble Kalman Filter for Geostatistical Applications*. Centre for Computational Geostatistics, Guidebook Series, Vol. 10, p.82.
- Zagayevskiy, Y. and Deutsch, C.V. 2011a. Temperature and Seismic Data Integration with the Ensemble Kalman Filter (EnKF). *Centre for Computational Geostatistics: Report 13*, 208-1 – 208-24.
- Zagayevskiy, Y. and Deutsch, C.V. 2011b. Temperature and Pressure Dependent Gassmann's Fluid Substitution Model for Generation of Synthetic Seismic Attributes. *Centre for Computational Geostatistics: Report 13*, 209-1 – 209-18.
- Zagayevskiy, Y. and Deutsch, C.V. 2011c. Application of Ensemble Kalman Filter (EnKF) to Realistic Oil Sands Data. *Centre for Computational Geostatistics: Report 13*, 210-1 – 210-8.

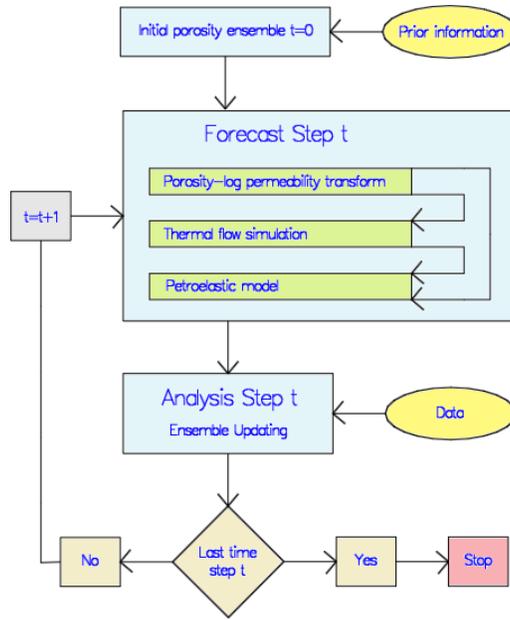


Figure 1: Data integration flow chart for SAGD petroleum reservoir characterization with the EnKF

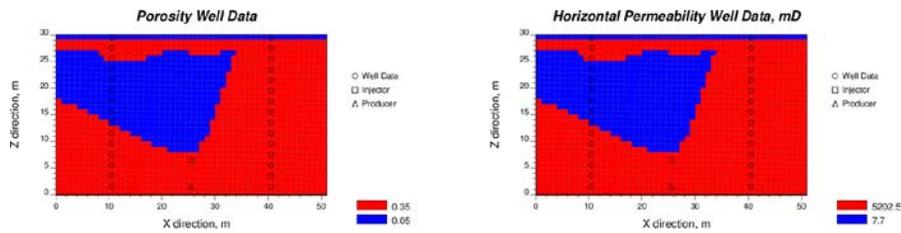


Figure 2: 2D porosity and horizontal permeability base case and well layout: vertical wells with observation locations, horizontal SAGD injector and producer

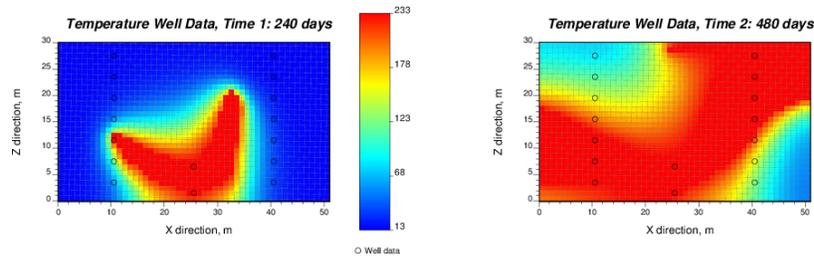


Figure 3: 2D temperature base case for 240 & 480 days after production begins with observation locations

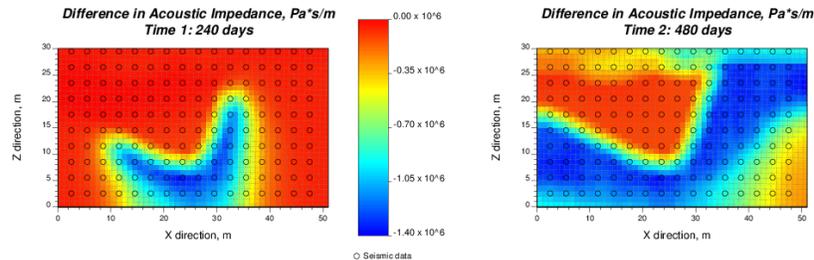


Figure 4: 2D difference of acoustic impedance base case for 240 & 480 days with observation location grid

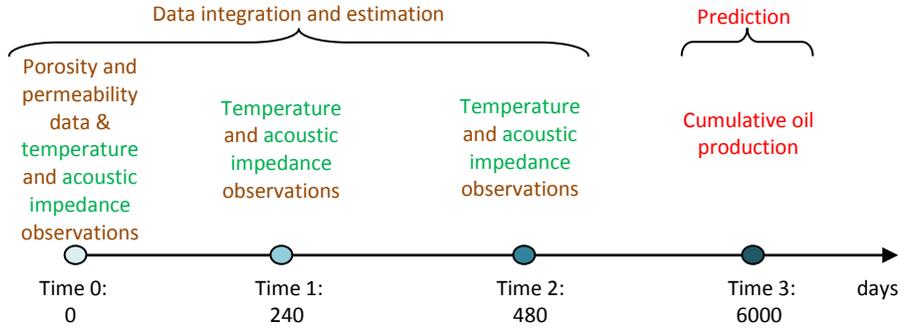


Figure 5: Timeline of data integration in 2D case study

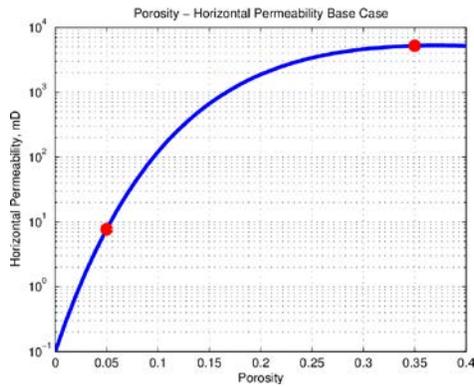


Figure 6: Porosity-horizontal permeability regression model: dots represent binary facies in 2D case study

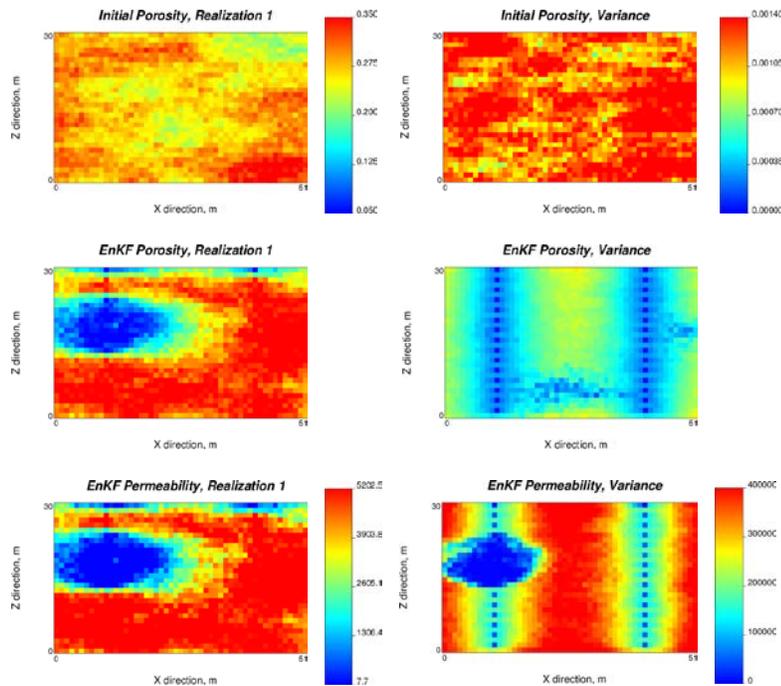


Figure 7: Evolution of porosity estimate, change of variance map, and evidence of strong porosity-permeability relationship. Low value artifacts on variance maps are introduced by limiting updated porosity estimate between 0.0 and 0.4.

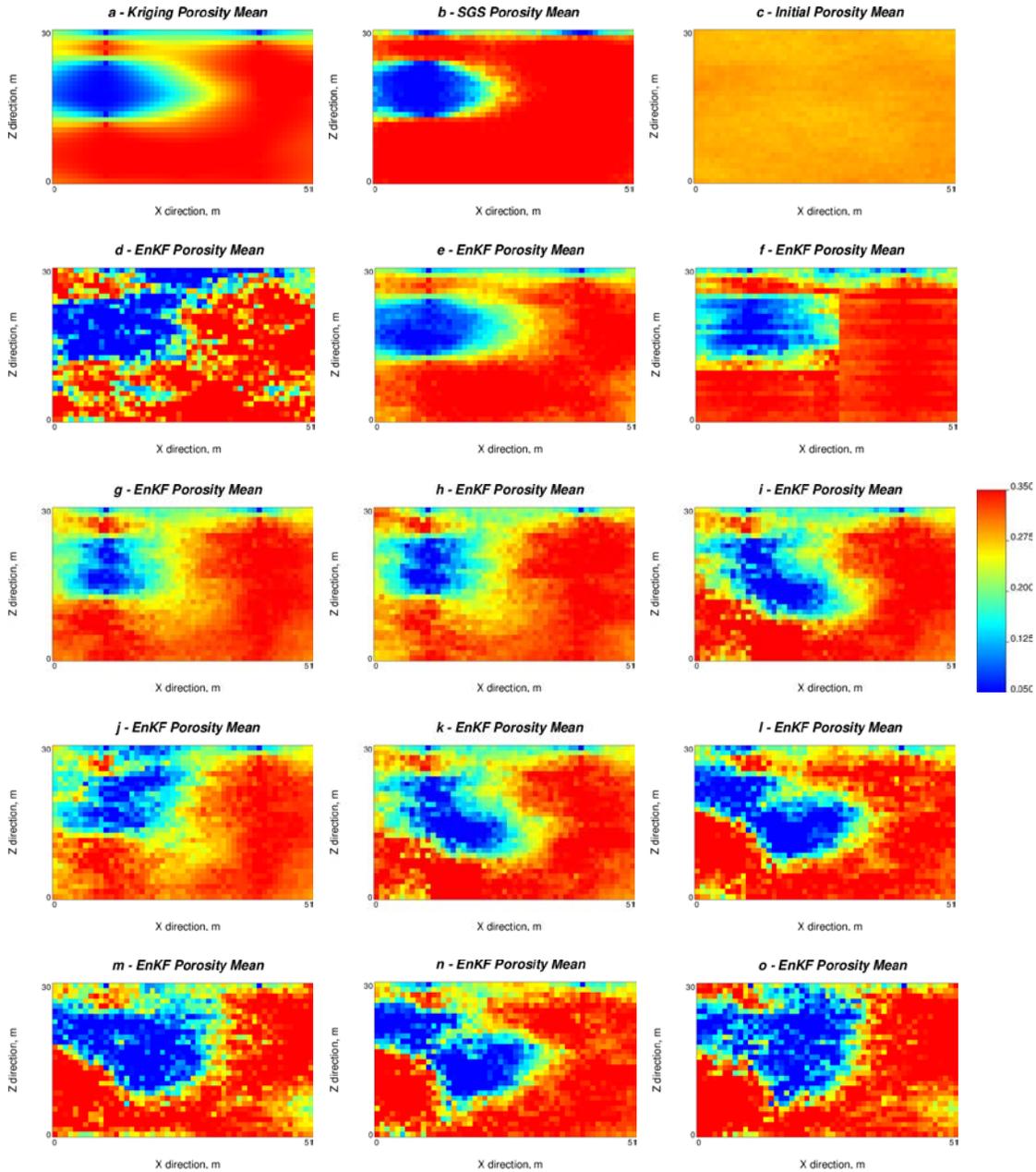


Figure 8: Mean value maps of porosity estimates from 50 realizations with localization of covariance matrix: a) kriging with porosity data, b) SGS with porosity data, c) initial ensemble, d) EnKF with porosity data, e) EnKF with porosity data, 1000 realizations and no localization, f) EnKF with porosity data and localization of updating matrix, g) EnKF with porosity data, h) EnKF with porosity and permeability data, i) EnKF with porosity data and temperature observations from day 240, j) EnKF with porosity data and temperature observations from day 240 with shortcut, k) EnKF with porosity data and temperature observations from day 240 and 480, l) EnKF with porosity data and difference in acoustic impedance observations from day 240, m) EnKF with porosity data and difference in acoustic impedance observations from day 240 and 480, n) EnKF with porosity data, temperature and difference in acoustic impedance observations from day 240, o) EnKF with porosity data, temperature and difference in acoustic impedance observations from day 240 and 480.

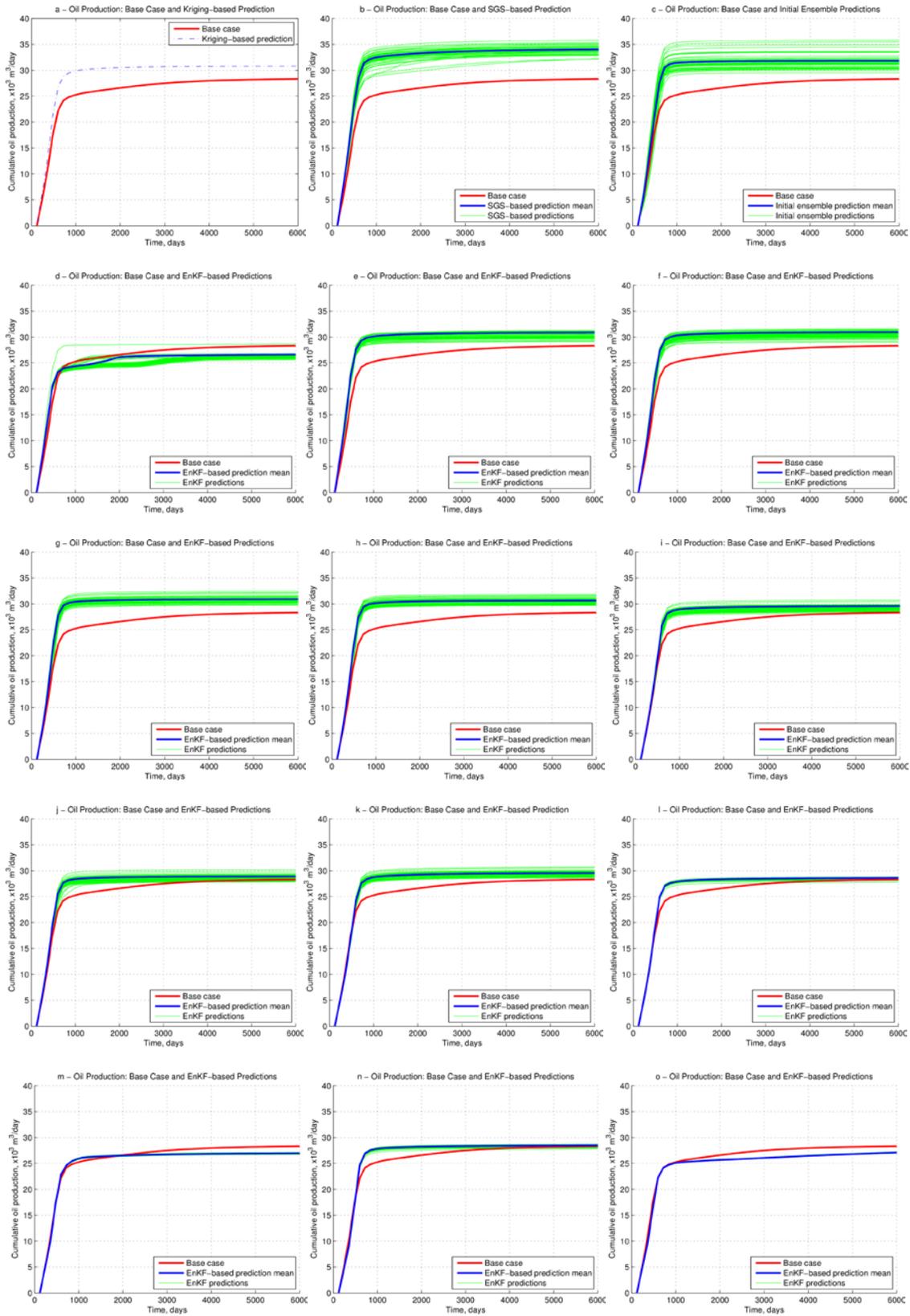


Figure 9: Prediction of cumulative oil production rate, 2D case study. See Figure 8 for description of cases.

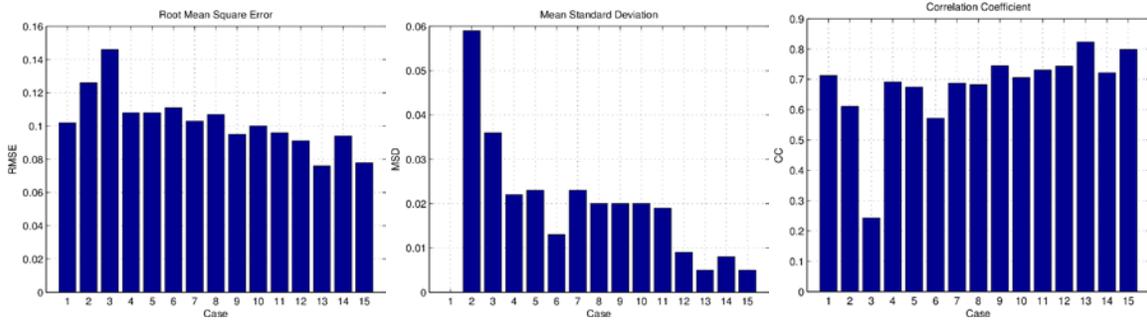


Figure 10: Results of validation calculations for 2D case study. The default parameters for the EnKF are 50 realizations with localized covariance matrix: 1) kriging with porosity data; 2) SGS with porosity data; 3) initial ensemble; 4) EnKF with porosity data; 5) EnKF with porosity data and 1000 realizations; 6) EnKF with porosity data and no localization; 7) EnKF with porosity data and localized updating matrix; 8) EnKF with porosity and permeability data; 9) EnKF with porosity data and temperature observations from time step 1; 10) EnKF with porosity data and temperature observations from time step 1 using shortcut; 11) EnKF with porosity data and temperature observations from time steps 1 and 2; 12) EnKF with porosity data and difference in acoustic impedance observations from time step 1; 13) EnKF with porosity data and difference in acoustic impedance observations from time steps 1 and 2; 14) EnKF with porosity data, temperature and difference in acoustic impedance observations from time step 1; 15) EnKF with porosity data, temperature and difference in acoustic impedance observations from time steps 1 and 2

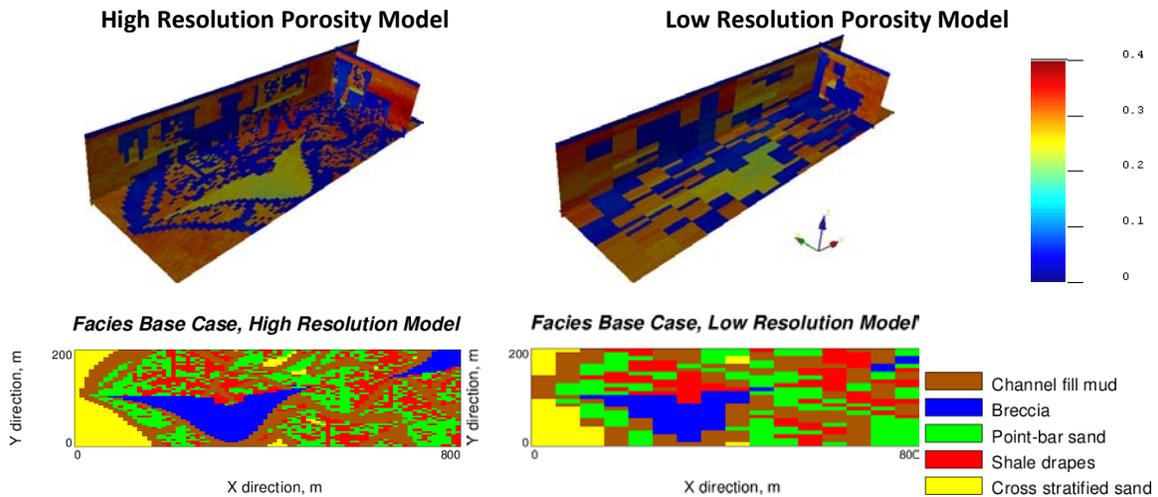


Figure 11: Facies base case for 3D case study: high resolution model for true production, low resolution model for the flow simulations. 3D and plan views are at 3 m elevation. Z axis is exaggerated by 5 times

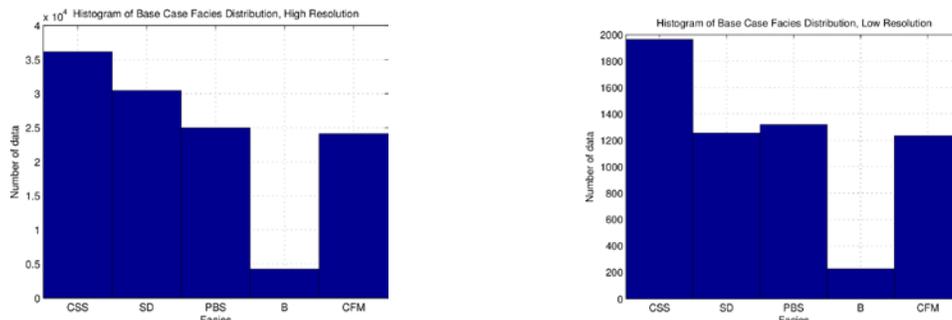


Figure 12: Histograms of facies proportions for high and low resolution models: cross-stratified sand (CSS), shale drapes (SD), point-bar sand (PBS), breccia (B), and channel fill mud (CFM)



Figure 13: Well configurations and SAGD layout for 3D case study

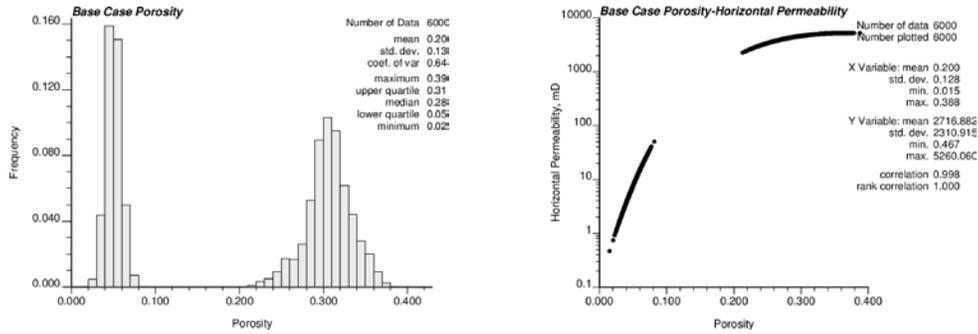


Figure 14: Histogram of base case porosity and base case porosity-horizontal permeability scatter plot

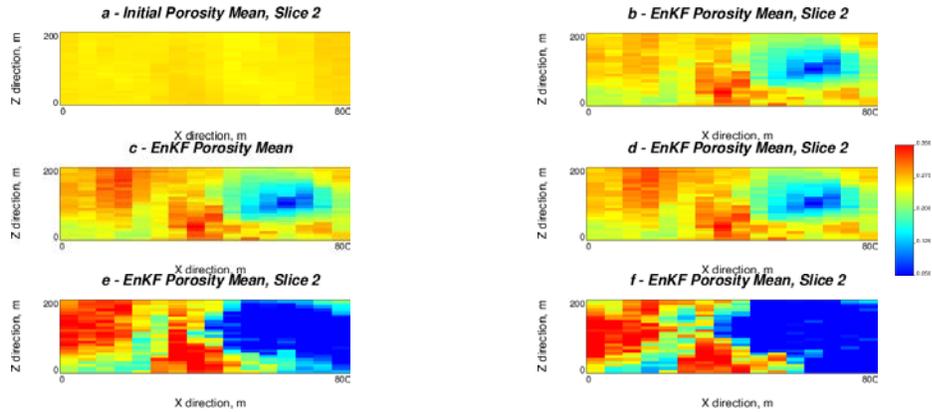


Figure 15: Plan view of means of EnKF porosity estimates derived from 50 realizations: a) initial Gaussian ensemble, b) porosity data, c) porosity data and temperature observations at 720 days, d) porosity data and temperature observations at 720 & 1440 days, e) porosity data and temperature observations at 720 days using shortcut, and f) porosity data and temperature observations at 720 & 1440 days using shortcut

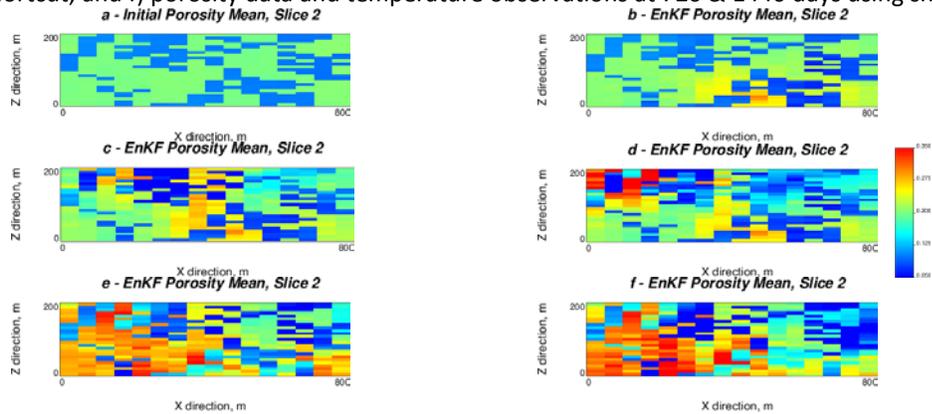


Figure 16: Plan views of means of EnKF porosity estimates derived from 50 realizations: a) initial binary ensemble, b) porosity data, c) porosity data and temperature observations at 720 days, d) porosity data and temperature observations at 720 & 1440 days, e) porosity data and temperature observations at 720 days using shortcut, and f) porosity data and temperature observations at 720 & 1440 days using shortcut

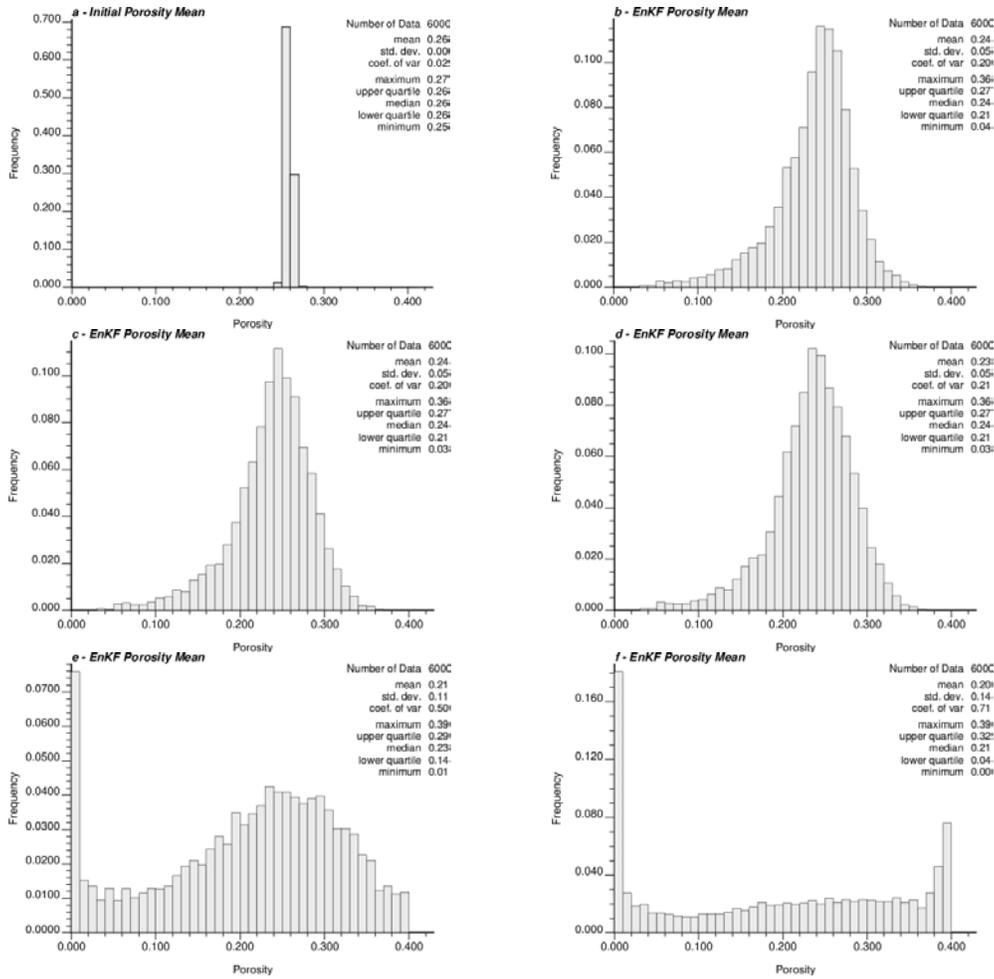


Figure 17: Histograms of EnKF porosity mean estimates derived from different cases using initial Gaussian ensemble, see Figure 15 for the details

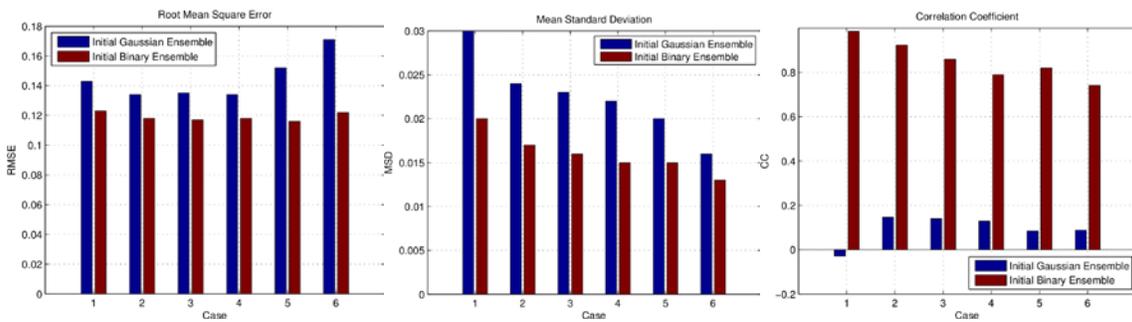


Figure 18: Results of validation calculations for 3D case study with initial Gaussian and binary ensembles. The default parameters for the EnKF are 50 realizations with localized covariance matrix: 1) initial ensemble; 2) EnKF with porosity data; 3) EnKF with porosity data and temperature observations from time step 1; 4) EnKF with porosity data and temperature observations from time steps 1 and 2; 5) EnKF with porosity data and temperature observations from time step 1 using shortcut; 6) EnKF with porosity data and temperature observations from time steps 1 and 2

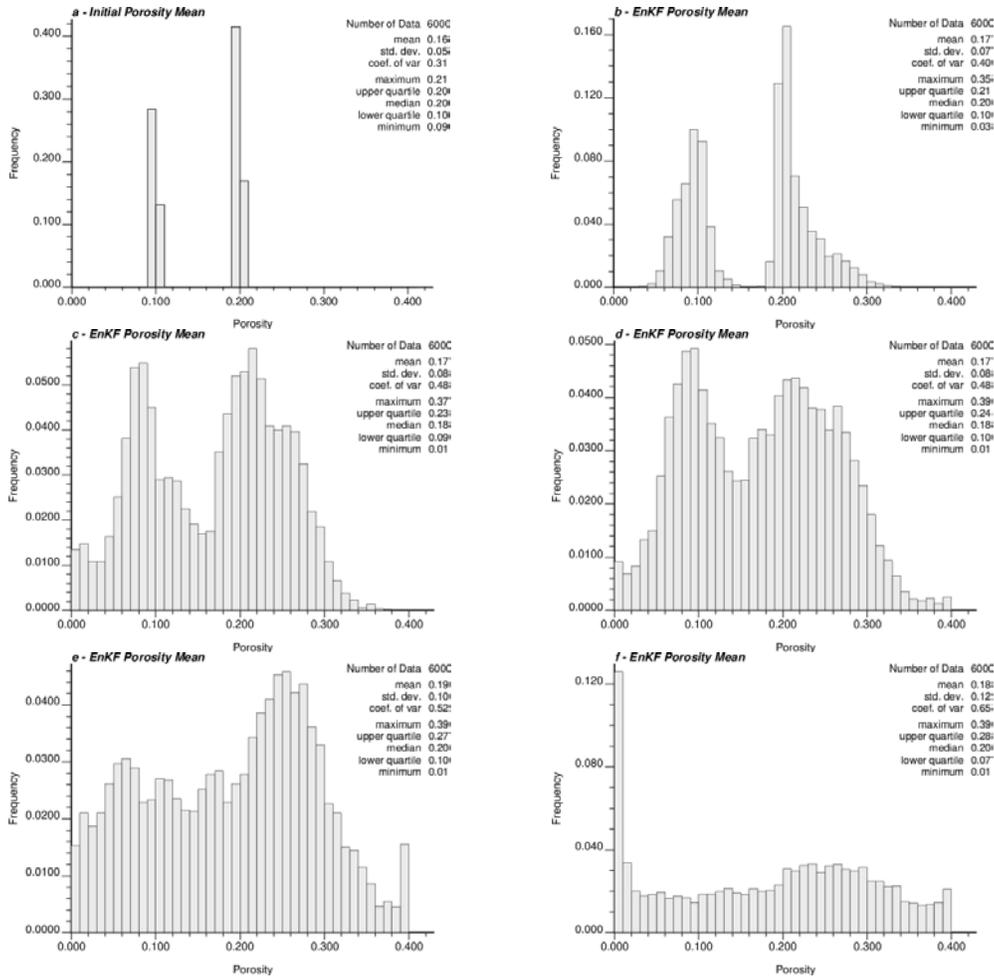


Figure 19: Histograms of EnKF porosity means estimates derived from different cases using initial binary ensemble, see Figure 15 for the details

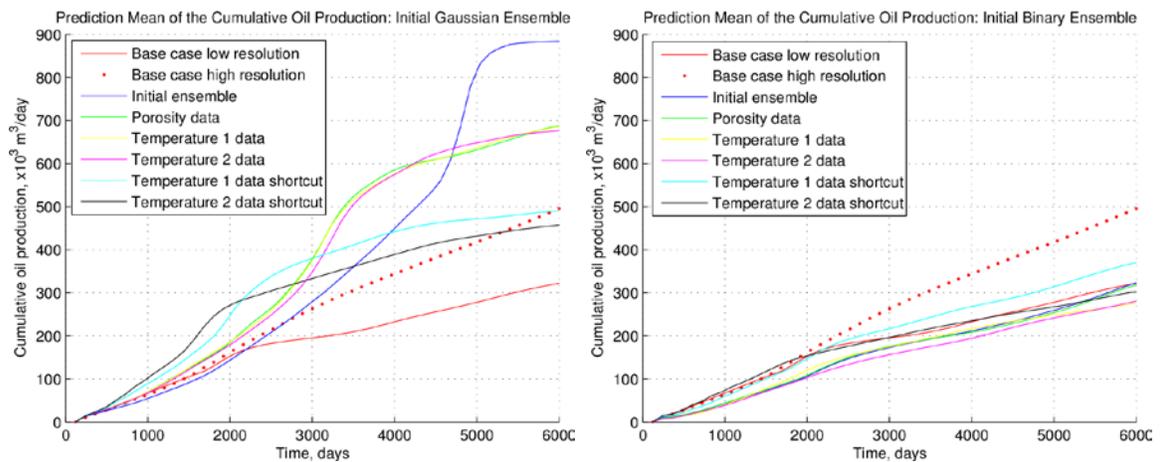


Figure 20: Cumulative oil production rate from the entire SAGD model: base case for low and high resolution 3D models, and prediction means of different cases with different initial ensembles