Document downloaded from:

http://hdl.handle.net/10251/73816

This paper must be cited as:

Li ., L.; Zhou ., H.; Gómez-Hernández, JJ. (2015). Two-point or multiple-point statistics? A comparison between the ensemble Kalman filtering and the ensemble pattern matching inverse methods. Advances in Water Resources. 86:297-310. doi:10.1016/j.advwatres.2015.05.014.

The final publication is available at

http://dx.doi.org/10.1016/j.advwatres.2015.05.014

Copyright Elsevier

Additional Information

Two-point or multiple-point statistics? A comparison between the ensemble Kalman filtering and the ensemble pattern matching inverse methods

Liangping Li[a,*], Sanjay Srinivasan[a], Haiyan Zhou[a], J. Jaime Gómez-Hernández[b]

[a] *Center for Petroleum and Geosystems Engineering Research, University of Texas at Austin,78712, Austin, USA*
[b] *Research Institute of Water and Environmental Engineering, Universitat Politècnica de València, 46022, Valencia, Spain*

Abstract

The Ensemble Kalman Filter (EnKF) has been commonly used to assimilate real time dynamic data into geologic models over the past decade. Despite its various advantages such as computational efficiency and its capability to handle multiple sources of uncertainty, the EnKF may not be used to reliably update models that are characterized by curvilinear geometries such as fluvial deposits where the permeable channels play a crucial role in the prediction of solute transport. It is well-known that the EnKF performs optimally for updating multi-Gaussian distributed fields, basically because it uses two-point statistics (i.e., covariances) to represent the relationship between the model parameters and between the model parameters and the observed response, and this is the only statistic necessary to fully characterize a multiGaussian distribution. The Ensemble PATtern matching (EnPAT) is an alternative ensemble based method that shows significant potential to condition complex geology such as channelized aquifers to dynamic data. The EnPAT is an evolution of the EnKF, replacing, in the analysis step, two-point statistics with multiple-point statistics. The advantages of EnPAT reside in its capability to honor the complex spatial connectivity of geologic structures as well as the measured static and dynamic data. In this work, the performance of the classical EnKF and the EnPAT are compared for modeling a synthetic channelized aquifer. The results reveal that the EnPAT yields better prediction of transport characteristics than the EnKF because it characterizes the conductivity heterogeneity better. Issues such as uncertainty of multiple variables and the effect of measurement errors on EnPAT results will be discussed.

Keywords: multiple-point statistics, rejection sampling, conditional simulation, ensemble Kalman filter, inverse method

*Corresponding author
Email address: liangpingli@utexas.edu (Liangping Li)

1. Introduction

Inverse methods have been used extensively in hydrology and petroleum engineering to identify spatial variations of geological parameters conditioned to observed dynamic data such as piezometric head and concentration, in order to improve flow and transport predictions. Inverse methods have evolved from manual trial-and-error approaches to real-time automatic data assimilation approaches; from deterministic estimation to stochastic simulation; from gradient-based minimization approaches to sampling-based approaches; and from multiGaussian-based methods to those without restrictive multiGaussian assumptions. An extensive description of the evolution of inverse methods and recent trends can be found in the work by Zhou et al. (2014).

The widely used ensemble Kalman filter (EnKF), an ensemble-based real-time data assimilation inverse method, was first proposed by Evensen (1994) as an extension of the extended Kalman filter. In the EnKF, the cross-correlations of the parameters and the state variables are explicitly calculated through an ensemble of realizations rather than approximated through a Taylor series expansion of the transfer function (e.g., Leng and Yeh, 2003). The ensemble Kalman filter has increasingly been used in multiple disciplines such as petroleum engineering and hydrogeology because of its computational efficiency and its real-time data assimilation capability (e.g., Wen and Chen, 2006; Gu and Oliver, 2006; Chen and Zhang, 2006; Hendricks Franssen and Kinzelbach, 2008; Bailey and Baù, 2010; Camporese et al., 2011; Li et al., 2011; Panzeri et al., 2013, 2014b,c,a; Gharamti and Hoteit, 2014). For instance, Chen and Zhang (2006) applied standard EnKF to a groundwater system in order to evaluate the sensitivity of inverted/updated parameters to factors such as ensemble size and frequency of conditioning data. Hendricks Franssen and Kinzelbach (2008) applied EnKF to a field case study and discussed the filter inbreeding issue in detail. Panzeri et al. (2013) coupled EnKF with moment equations to circumvent the computational cost needed in the Monte Carlo simulations, and applied this novel approach in a real case study (Panzeri et al., 2014c). Gharamti et al. (2014b) proposed a hybrid formulation of the EnKF and optimal interpolation that integrates both the ensemble sample covariance and a static background covariance in order to reduce the size of the ensemble and to avoid filter divergence. Panzeri et al. (2014a) developed a two-step updating scheme to integrate dynamic data into reservoir models exhibiting complex geology; specifically, the geometry of facies is first handled using a Markov mesh model, and then the EnKF is applied to calibrate the conductivities within each facies.

In fluvial depositions and fractured systems, hydraulic conductivity is commonly assumed to follow a multi-modal probability distribution. In other words, spatial variation in conductivity can be thought of as outcome from several random processes that characterize the geo-material (i.e., facies). Winter et al.

(2002) and Winter et al. (2003) discussed this type of "composite medium" in detail, and, for example, Guadagnini et al. (2003) and Riva et al. (2008) have applied this concept for aquifer modeling. Multiple-point statistics (MPS) methods are becoming popular for characterizing fluvial depositions and fractured aquifers. MPS uses (only) the observed static data (i.e., measured conductivity data) for conditioning geologic models. Compared to traditional two-point covariance-based geostatistical methods, MPS has the capability to effectively reproduce complex structures observed in a conceptual model (i.e., training image). Several MPS algorithms have been described in the literature since SNESIM, the first MPS code, was developed by Strebelle (2002). Hu and Chugunova (2008) presented a comprehensive review of MPS methods.

The challenge is to integrate dynamic data into the MPS-based geological modeling procedures. More specifically, the key question is how to condition non-multiGaussian fields to dynamic data that are related to the model parameters through highly non-linear relationships. Recently, a number of authors have tried to apply the EnKF to an ensemble of MPS-based non-multi-Gaussian conductivity fields where the uncertainty is mainly due to the spatial distribution of geologic facies. However, the fact that the analysis equations in EnKF are equivalent to the normal equations (or cokriging equations) implies that the EnKF is optimal for multi-Gaussian fields and linear state equations (Aanonsen et al., 2009). In other words, using only two-point covariances and cross-covariances between parameters and state variables in the analysis step of the EnKF, the heterogeneity features that are controlled by higher-order statistics may not be preserved during the updating process. For this reason, a number of variations to the EnKF-based methods have been proposed to ensure that the connectivity prescribed by MPS simulations is preserved. For instance, Jafarpour and Khodabakhshi (2011) introduced a probability conditioning method, in which a probability field (i.e., ensemble mean of the indicator values of conductivity) is first derived by assimilating the dynamic data, and then the MPS conductivity realizations are regenerated using the calculated probability field as soft data. Sun et al. (2009) and Dovera and Della Rossa (2011) proposed to couple mixture Gaussian models and the EnKF to preserve the spatial structure of MPS conductivity simulations. Sarma and Chen (2009) introduced a kernel EnKF approach applied to MPS conductivity simulations. Zhou et al. (2011) and Li et al. (2012b) developed a normal score EnKF (NS-EnKF) approach in which a normal-score transformation is applied to both the non-Gaussian parameters and state variables prior to the analysis step. Hu et al. (2013) proposed to update the uniform random numbers that are used to draw the conductivity values from local conditional probability in the context of a sequential MPS simulation (as implemented in SNESIM, for example). Ping and Zhang (2013) presented a vector-based level-set parameterization approach for channelized aquifers, and then combined it with the EnKF to match the observed dynamic data. All of the

above mentioned EnKF-based methods accomplish the goal of reproducing non-Gaussian reservoir models to varying degrees of success but they may still result in suboptimal solutions because the analysis step is still based on two-point covariances and cross-covariances.

Unlike the previous variants of the EnKF, Zhou et al. (2012) proposed a fully non-Gaussian stochastic inverse method, termed the Ensemble PATtern matching method (EnPAT), which is an evolution of the EnKF to deal with the issue of reproduction of spatial patterns prescribed by MPS simulations. In EnPAT, the correlation between model parameter and state variables is delineated by MPS (i.e., pattern) rather than by traditional two-point covariances, and thus curvilinear heterogeneities can be preserved while the dynamic data are integrated. Li et al. (2014) further extended this method to simultaneously estimate parameter and state variables so that a better characterization at multiple scales is achieved. To improve the computational efficiency, Li et al. (2013) coupled the EnPAT algorithm with a pilot-point scheme such as in the implementation of the self-calibration inverse method (Gómez-Hernández et al., 1997; Wen et al., 1999).

In this work, we highlight the capabilities of the EnPAT method to assimilate dynamic data by comparing its performance to the standard EnKF. First, the EnPAT is extended to handle continuous conductivity fields. Then, the performance of EnKF and EnPAT is compared on a synthetic aquifer example that is characterized by curvilinear channels with high permeability. Also, in order to explore the space of posterior uncertainty, Bayes' rejection sampling method is applied in a benchmark case. The performance of EnPAT is evaluated in terms of aquifer characterization, and flow and transport predictions. Finally, we discuss the advantages and drawbacks of the EnPAT method.

The paper will continue as follows: the EnKF and EnPAT algorithms are described in section 2; in section 3, a synthetic example is analyzed using both the EnKF and the EnPAT methods. There is a discussion of the main results in section 4; the paper ends with a summary.

2. The EnKF and EnPAT algorithms

2.1. General framework

The main procedure of both algorithms includes two steps: forecast and analysis. The difference between the EnKF and the EnPAT resides in the analysis step, the EnKF is based on two-point covariances and the EnPAT on multiple-point statistics. The specifics are as follows:

1. Initialization step

Generate a set of initial models conditioned to the measured static data. In complex geological formations such as fluvial deposits, an MPS simulation method is commonly employed, which uses a conceptual model represented by a training image. Examples of MPS algorithms are the single normal equation simulation (SNESIM) (Strebelle, 2002) and the direct sampling method (DS) (Mariethoz et al., 2010b). Note that the initial conductivity realizations are the same for both the comparison of the EnKF and EnPAT methods in this paper.

2. Forecast step

   For each conductivity realization $\mathbf{X}$, the groundwater flow equation is solved from time $t = 0$ to $t = k$, i.e.,

   $$\mathbf{Y}_k = f(\mathbf{X}_{k-1}) \tag{1}$$

   where f represents the groundwater flow model, boundary conditions as well as sources and sinks. $\mathbf{Y}_k$ denotes the simulated piezometric head at time $t = k$. The conductivity $\mathbf{X}_{k-1}$ and corresponding head $\mathbf{Y}_k$ will be used in the analysis step to derive an updated conductivity $\mathbf{X}_k$.

3. Analysis step

   Given the mismatch between the observed state $\mathbf{Y}_k^{obs}$ and the forecasted state values, the ensemble of conductivity $\mathbf{X}$ is updated from time $t = k - 1$ to time $t = k$. Specific analysis schemes for each method will be discussed in subsequent subsections for the EnKF and EnPAT.

4. Loop back to step 2 for the next time step. The forecast and analysis loop starts again with the updated conductivity $\mathbf{X}_k$ as the new parameters in equation (1). The loop ends when all observation data have been integrated.

Both methods are based on utilizing the ensemble of realizations and the corresponding responses computed using a nonlinear model (e.g., physical model) to infer the statistics needed for the analysis step. The resultant updated ensemble provides a quantification of the residual uncertainty. We demonstrate the method in the framework of groundwater modeling, but it is very easy to replace the groundwater flow equation in (1) by any other state equation, so that both methods could be applied to a variety of data conditioning problems, i.e. multiphase and compositional flow simulation in the context of petroleum engineering.

We also note that both the spatially varying rock conductivity and piezometric head are updated in the analysis step in the original implementation of EnKF/EnPAT. However, we only retain the updated conductivities for the next time step data assimilation because the updated piezometric head realizations might not be fully consistent with the updated conductivities in terms of satisfying mass balance constraints.

For this reason, after updating the conductivities at each time step, equation (1) is rerun from time $t = 0$ to ensure that the piezometric heads used for the analysis at the next time step are fully consistent with the current conductivities. This rerun procedure significantly increases the computational cost of the filter. There has been some debate regarding the use of the updated piezometric heads obtained after the analysis step for the the subsequent update steps (e.g., Wen and Chen, 2006; Hendricks Franssen and Kinzelbach, 2008; Schöniger et al., 2012; Jafarpour and Khodabakhshi, 2011), however, for data integration in strongly heterogeneous systems it is quite possible that statistical updating of dynamic variables such as head and saturations may result in serious mass balance errors.

2.2. Analysis step for the EnKF

An extensive description of the EnKF algorithm can be found in Burgers et al. (1998) and Evensen (2003). The algorithm for the analysis step is as follows:

(a) Build the augmented vector $\mathbf{\Psi}_k$, which includes both the conductivity $\mathbf{X}_k$ and piezometric head $\mathbf{Y}_k$. It can be expressed as,

$$\mathbf{\Psi}_{k,j} = \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix}_{k,j} \tag{2}$$

where $\mathbf{\Psi}_{k,j}$ denotes the jth ensemble member of the augmented vector at time $t = k$.

(b) Calculate the Kalman gain ($\mathbf{G}_k$),

$$\mathbf{G}_k = \mathbf{P}_k^f \mathbf{H}^T \left(\mathbf{H} \mathbf{P}_k^f \mathbf{H}^T + \mathbf{R}_k \right)^{-1}, \tag{3}$$

where the superscript f means forecast; $\mathbf{H}$ is an observation matrix generally composed of $0's$ and $1's$; $\mathbf{R}_k$ is the measurement error covariance matrix, assumed diagonal (i.e., measurement errors between well locations are independent); $\mathbf{P}_k^f$ is the covariance matrix of the augmented vector $\mathbf{\Psi}_k$, which can be estimated from the ensemble as,

$$\mathbf{P}_k^f \approx \frac{1}{N_e - 1} \sum_{j=1}^{N_e} \left(\mathbf{\Psi}_k^f - \overline{\mathbf{\Psi}}_k^f \right) \left(\mathbf{\Psi}_k^f - \overline{\mathbf{\Psi}}_k^f \right)^T \tag{4}$$

$$\overline{\mathbf{\Psi}}_k^f \approx \frac{1}{N_e} \sum_{j=1}^{N_e} \mathbf{\Psi}_{k,j}^f \tag{5}$$

where N_e is the ensemble size and $\overline{\mathbf{\Psi}}$ denotes ensemble mean.

Note that, in the numerical implementation, it is not necessary to calculate the full ensemble covariance $\mathbf{P}_k^f$ since we can compute the terms $\mathbf{P}_k^f\mathbf{H}^T$ and $\mathbf{H}\mathbf{P}_k^f\mathbf{H}^T$ directly taking the advantage of the sparsity of matrix $\mathbf{H}$.

(c) Update the augmented vector $\mathbf{\Psi}_k$ by assimilating the observation data $\mathbf{Y}_k^{obs}$.

$$\mathbf{\Psi}_{k,j}^a = \mathbf{\Psi}_{k,j}^f + \mathbf{G}_k\Big(\mathbf{Y}_k^{obs} + \boldsymbol{\epsilon}_j - \mathbf{H}\mathbf{\Psi}_{k,j}^f\Big), \tag{6}$$

where superscript a denotes analysis and $\boldsymbol{\epsilon}_j$ represents a random observation error vector.

The update of the augmented vector depends on the Kalman gain, which, in turn, depends only on the sample covariance derived from the ensemble of realizations. Only two-point covariances are used, these covariances would completely characterize the multivariate distribution of the augmented state variables if they follow a multiGaussian distribution. The two-point covariances are insufficient to describe a non-Gaussian multivariate distribution. Besides, an intrinsic problem of the EnKF is that the experimental covariances computed from (4) are only an approximation of the "true" covariance unless large ensembles are used. Some approaches have already been implemented to alleviate this covariance inference issues such as localization and hybrid formulation (e.g., Hamill et al., 2001; Gharamti et al., 2014b).

The entire augmented vector, that is, both conductivity and piezometric heads, is updated during the analysis step. The computational cost of the analysis step is small as compared with the forecast step. Gharamti et al. (2014a) discussed the computational cost of each EnKF step in detail. Moreover, as mentioned before, the resulting updated piezometric heads may not be consistent with the updated conductivity values, in the sense that they might violate the physics of the flow equation, and for this reason we choose to rerun the forecast step from time zero with the latest updated conductivities.

The EnKF is designed for updating continuous variables, although some authors have proposed some types of reparameterization of categorical variables (Wen et al., 2000; Hu et al., 2013) which can be used in EnKF. Note also that the updating Equation (6) does not take into account the underlying physics, and therefore, updating some variables, such as porosity, can result, directly, in non-physical values.

2.3. Analysis step for the EnPAT

The EnPAT method was first proposed by Zhou et al. (2012) and further improved by Li et al. (2013). Figure 1 shows the flowchart of the EnPAT algorithm. The implementation of the analysis step in the EnPAT algorithm is as follows.

(a) Start the loop for the estimation of conductivity for realization number 1, $\mathbf{X}_{k,j}$ $(t = k; j = 1)$

(b) Define a set of randomly distributed pilot points $P_i(X), i = 1, 2, \cdots, N_p$, where the number of pilot points, N_p, is defined by the user.

(c) Start the loop to estimate both the conductivity x_i and piezometric head y_i at pilot point i; if the location of ith pilot point coincides with measurement, go to step (d); otherwise,

(I) Build the spatial pattern $\psi_{k,i,j}$ of both conductivity as well as piezometric head data as shown in Figure 2A.

$$\psi_{k,i,j} = \left[\begin{array}{c} \mathbf{x} \\ \mathbf{y} \end{array} \right]_{k,i,j} \tag{7}$$

where $\psi_{k,i,j}$ represents the joint pattern of conductivity and piezometric head values associated with the ith node to be estimated, for the jth member of ensemble and at time $t = k$. $\mathbf{x}$ and $\mathbf{y}$ denote conductivity and piezometric head data, respectively, and they can be either measured data or previously simulated values. The size of the joint pattern $\psi_{k,i,j}$ depends on a user-predefined maximum search radius and maximum number of conditioning data. The conditioning data are searched following a spiral path away from the node to be updated.

(II) Look for a matching pattern in the ensemble of joint realizations of conductivity and piezometric head. This search is performed by visiting each pair of realizations (conductivity and head) according to a random path $P_\lambda, \lambda = 1, 2, \cdots, N_r$, where N_r is the number of realizations in the ensemble.

(III) Start the loop with realization $\lambda = 1$:

i. Locate the candidate pattern $\overline{\psi}_{k,i,\lambda}$ in the pair of training images for conductivity and piezometric head. The candidate pattern $\overline{\psi}_{k,i,\lambda}$ is fixed at location i since piezometric head is affected by boundary conditions and source terms.

$$\overline{\psi}_{k,i,\lambda} = \left[\begin{array}{c} \overline{\mathbf{x}} \\ \overline{\mathbf{y}} \end{array} \right]_{k,i,\lambda} \tag{8}$$

where $\overline{\mathbf{x}}$ and $\overline{\mathbf{y}}$ represent the last conductivity estimate (from the analysis step at time $t = k-1$) and the corresponding simulated head for time $t = k$, respectively.

ii. Calculate the distance $d_{(\psi,\overline{\psi})}$ between the conditional pattern $(\psi_{k,i,j})$ and the candidate pattern $(\overline{\psi}_{k,i,\lambda})$. Because the pattern is composed of both conductivity and piezometric head, in order to avoid specifying a weight to be assigned for each variable, distance is calculated for

the conductivity $d_{(\mathbf{x},\overline{\mathbf{x}})}$ and head $d_{(\mathbf{y},\overline{\mathbf{y}})}$, separately and then each distance is tested against tolerance values as explained below. Zhou et al. (2012) discussed, in details, several possible distance functions: Specifically, the Manhattan distance is best for categorical variables while the weighted Euclidean distance is preferred for continuous variables.

iii. Compare the calculated distance values with predefined tolerance values (ξ_x, ξ_y). If both distances for conductivity and head are lower than their tolerances ($d^{\lambda}_{(\mathbf{x},\overline{\mathbf{x}})} < \xi_{\mathbf{x}}$ and $d^{\lambda}_{(\mathbf{y},\overline{\mathbf{y}})} < \xi_{\mathbf{y}}$), copy the conductivity and piezometric head values at node i from the joint realizations λ to the node being estimated i ($x_{i,j} = x_{i,\lambda}$, $y_{i,j} = y_{i,\lambda}$) and go to step (d).

iv. Otherwise, go to next realization looking for a better candidate pattern (i.e., step (i)) and set $\lambda = \lambda + 1$.

(IV) When none of the candidate patterns meets the tolerance restrictions, chose the ϑth realization having the smallest distances for both conductivity and piezometric head ($d^{\vartheta}_{(\mathbf{x},\overline{\mathbf{x}})} = d_{\mathbf{x},min}$ and $d^{\vartheta}_{(\mathbf{y},\overline{\mathbf{y}})} = d_{\mathbf{y},min}$), and then copy the conductivity and head from this realization onto the node being estimated i ($x_{i,j} = x_{i,\vartheta}$, $y_{i,j} = y_{i,\vartheta}$).

(d) Go to the next pilot point (i.e., step (b)) and set $i = i + 1$.

(e) Complete the realization by using the pilot point conductivities generated as conditioning data and employing a traditional MPS method such as DS to perform spatial interpolation.

(f) Go to the next conductivity realization (i.e., step (a)) and set $j = j + 1$.

In the EnPAT, unlike in the EnKF, during the analysis step, conductivities and piezometric heads are sequentially generated, node by node, and realization by realization. This pixelwise simulation therefore follows the characteristics of traditional sequential simulation geostatistical methods. Actually, the EnPAT method is developed based on the direct sampling (DS) MPS method (Mariethoz et al., 2010b) by adding the state variable (piezometric head, here) into the parameter pattern and by using the ensemble pairs of conductivity and piezometric joint training images to search for matching spatial pattern instead of using a single training image as in DS. It is the use of the ensemble of joint training images, and the location-specific search through the ensemble for suitable candidate patterns that differentiates this method from others and renders it possible to condition this method to dynamic data (see Figure 2B). To keep parameter and state variables balanced in the patterns, conductivity and piezometric head are both estimated at the pilot point locations (Li et al., 2014). Then, pilot point values are used as conditioning data, and the traditional Direct Sampling MPS method is used to extrapolate the pilot point values to generate conductivities at the remaining unsampled locations. This dual approach, that is, first generate the conductivities at the

pilot points using an ensemble of training images, and filling in the gaps by standard MPS, reduces the computational cost without unduly impacting the accuracy of the generated models. Li et al. (2013) analyzed the impact of the number of pilot point on the computational cost. The larger the number of pilot points, the better the accuracy of predictions made using the models but the higher the associated computational cost of generating the models.

The EnPAT can be applied to both categorical and continuous variables. The only difference resides in the specification of the distance function used to quantify the dissimilarity between conditioning pattern and the candidate patterns. Li et al. (2013) demonstrates the application of EnPAT to categorical conductivity fields.

3. An illustrative example

3.1. Reference Field

A single-phase transient groundwater flow example is presented to compare the effectiveness of the EnKF and the EnPAT for dynamic data integration when considering non-multiGaussian conductivity fields. The aquifer has $50 \times 50 \times 1$ cells of size 1 m $\times$ 1 m $\times$ 1 m. The east and west sides of the aquifer are constant head boundaries with prescribed head values of 0 m, the other two sides are no flow boundaries. Porosity and specific storage are assumed to be constant and set equal to 0.3 and 0.02 m^{-1}, respectively. An injection well (#5) is located at the center of the aquifer with a constant injection rate $Q = 25$ m^3/d (see Figure 3). The total simulation time is 30 days, discretized into 10 time steps with varying time step size following a geometric sequence of ratio 1.2. The head data collected from 9 wells in the first five time steps will be used for conditioning.

Several MPS algorithms presented in the literature have the capability to handle continuous variables such as hydraulic conductivity. Here, the DS method is used to generate the reference conductivity field. The training image is obtained by superimposing a continuous conductivity field generated by sequential Gaussian simulation (Gómez-Hernández and Journel, 1993; Gómez-Hernández and Cassiraga, 1994) on the facies model in Strebelle (2002) (see Figure 3A). The distance function plays a key role in DS and EnPAT. Because the conductivity and piezometric head are both continuous variables, the weighted Euclidian distance function is used to calculate the distance between the patterns, and it is defined as

$$d_{(\psi,\overline{\psi})} = \left[\frac{1}{\sum_{i=1}^{n} h_i^{-1}} \sum_{i=1}^{n} h_i^{-1} \frac{(\psi(x_i) - \overline{\psi}(x_i))^2}{d_{max}^2} \right]^{1/2} \quad d \in [0,1] \tag{9}$$

where n is the number of elements in the pattern; x_n represents the elements in the pattern, which can be conductivity or head values; $\psi(x_n)$ and $\overline{\psi}(x_n)$ are the conditional pattern and candidate pattern, respectively; h_i is the Euclidean distance between element i and the element being simulated, d_{max} is the maximum absolute difference between the conditional pattern and the candidate pattern. As already mentioned, in order to avoid specifying the weight of each variable in the joint pattern, we calculate the distances for the conductivity and piezometric head, separately, and thus two distance tolerance values have to be predefined to decide whether a candidate joint pattern is accepted or not.

The set of parameters used in DS and EnPAT such as search radius, number of conditioning data and distance tolerances, are listed in Table 1. To understand the specific role of these parameters in the DS, the reader is referred to the work by Mariethoz et al. (2010b). Also, Meerschman et al. (2012) conducted an extensive sensitivity analysis of those parameters. Figure 3B shows the reference conductivity field generated by DS, which will be used as the "true" conductivity field. The conductivities collected from 9 wells will be used as the observed conductivity data (i.e., hard data). The spatial locations of the 9 wells can be found in Figure 3B.

Five hundred conductivity realizations, conditioned to the measured conductivity data, are generated by DS using the same training image as in the reference. The EnKF and the EnPAT are used to update the prior models by conditioning to the piezometric head data computed on the reference model. Table 1 summarizes the parameters used and flow configurations of the synthetic example.

The major goal of this paper is to compare the performance of the EnKF and the EnPAT to integrate dynamic data on a channelized aquifer. In order to check the posterior uncertainty after data integration, the Bayesian rejection sampling is included in the comparison as well, and described in the next section.

3.2. Rejection Sampling

The Bayesian rejection sampling has been implemented in the studies by Mariethoz et al. (2010a), Scheidt et al. (2014) and Satija and Caers (2015) to explore the posterior uncertainty after the integration of dynamic data. Unlike the EnKF that is based on model updating and the EnPAT that is based on model generation using training sets of conductivity and piezometric head, rejection sampling is based on sampling a subset of prior models that exhibit flow characteristics most similar to the observed dynamic data. Its implementation steps are briefly introduced as follows,

(a) Generate a prior model (m) using, for instance, a geostatistical method;

(b) Run a physical model (i.e., MODFLOW) to get the simulated heads;

(c) Evaluate an objective function $O(m)$. Here, the objective function is defined as the weighted square difference between the simulated and observed head data.

$$O(m) = \frac{1}{N_{well} \times N_{times}} \sum_{w=1}^{N_{wells}} \sum_{k=1}^{N_{times}} \left(\mathbf{y}_{obs}^{k,w} - \mathbf{y}^{k,w}(m) \right)^2 \tag{10}$$

where N_{well} and N_{times} indicate the number of wells and times in the data conditioning.

(d) Evaluate a likelihood $L(m)$. We assume the likelihood function to be Gaussian.

$$L(m) = K_l e^{-\frac{O(m)}{2\sigma^2}} \tag{11}$$

where K_l is a scaling factor such that the maximum $L(m)$ value is 1; σ is related to the measurement error of head.

(e) Sample a random number u from a uniform distribution over (0, 1);

(f) Accept the model m if $u < L(m)$.

Repeat the above procedure until an adequate number of models are sampled that could be used to represent the posterior uncertainty. As we see, the result of rejection sampling method is dependent on the definition of the objective and likelihood functions and the measurement error σ. In the benchmarking case, 114 models are sampled by evaluating 10000 forward simulations. The rejection sampling method is prohibitively expensive and quite likely inapplicable for most practical applications.

3.3. Results

3.3.1. Hydraulic Conductivity Characterization

Figure 4 shows the histogram of logconductivity for the reference field, and the histograms computed on the ensemble of conductivity fields after assimilating piezometric head data corresponding to the first five time steps for EnKF, EnPAT and rejection sampling. For the ensemble obtained by the EnKF, the histogram loses its bimodal characteristic and becomes more Gaussian-like (results which are consistent with those by Zhou et al. (2011)). On the contrary, EnPAT preserves the bimodal features of the reference field. In essence, the EnPAT preserves not only the lower-order moments (i.e., the mean and variance) but also the complex spatial structure of the conductivity field described by the multiple-point statistics. It is evident that rejection sampling preserves both the lower-order moments of the spatial distribution as well as the higher-order moments because a subset of the prior models is retained after evaluating the likelihood.

Figure 5 displays two individual realizations before and after conditioning to head data using the EnKF, the EnPAT and rejection sampling. The updated realizations using the EnKF tend to lose the channel features of the training image. Partly, this is expected because in the EnKF each updated realization can be interpreted as a linear combination of the ensemble of prior models. Another reason is that, in the long run, the EnKF realizations tend to lose the bimodality exhibited by the channelized prior model. For the EnPAT and rejection sampling, the individual conductivity realizations retain the channel features of the training image, and end with similar channel structures as the ones in the reference field because of the conditioning on piezometric head.

Figure 6 shows the ensemble mean and variance of logconductivity before and after conditioning to piezometric head data using the EnKF, the EnPAT and rejection sampling. In all cases, the mean maps of logconductivity patterns capture the main patterns of heterogeneity after head conditioning. Recall that the prior geological structures are lost in the updated models obtained using the EnKF. Looking at the ensemble variance maps, the one obtained by the EnKF has quite large values overall except in the vicinity of hard conditioning data, whereas the ensemble obtained by the EnPAT has low ensemble variance overall with only some uncertainty remaining at the location of the channel borders in the reference. Comparing the EnPAT with rejection sampling, both cases could identify the location of the channels, while the EnPAT yields smaller posterior uncertainty as shown in the variance maps. This is because the EnPAT is a multiple-point statistics-based method in which the ensemble pattern-searching procedure would require the ensemble size to be large enough in order to reproduce the characteristics of the "true" multivariate distribution. Small ensemble size could lead to overestimation of the posterior uncertainty and may lead to filter collapse as commonly observed in EnKF.

3.3.2. Flow and Transport Predictions

Essentially, we perform assimilation of dynamic data in order to improve the accuracy of our forecasts. The updated models are expected to have a better prediction performance since they are conditioned to different types of data (here, both the measured conductivities and piezometric head data). We have performed a flow and transport simulation to check how the updated models obtained by the EnKF and the EnPAT compare.

For the flow simulation, the configuration of wells is kept the same as before. The flow simulation is rerun from time zero until 30 days using the updated models. Figure 7 displays the piezometric head predictions for two wells, using the prior models and the updated models obtained by the EnKF and the EnPAT. As we can see, in both cases, the mean of the ensemble of heads is close to the reference head and the spread

of heads about the mean diminishes after conditioning to the first five time steps of head data. Also, the performance of the EnKF and the EnPAT is similar in this example, even if the individual realizations of the EnKF can not preserve the prior geological structures. This maybe because flow reaches almost steady-state after the first five-time steps and thus the forecast is almost unchanged.

For the transport simulation, using the same flow configuration as before, a conservative tracer with a concentration of 10 ppm is continuously injected at well #5. Figure 8 shows the concentration after 30 days, for the reference field, and the ensemble mean and variance of the predicted concentrations for the prior models and for the updated models using the EnKF and the EnPAT. Because no tracer data are used for conditioning, prediction of concentration profile would be a good check on the effectiveness of the algorithms for data conditioning and parameter estimation, particularly to see how effectively they have captured the effect of the channels in the transport simulations (Gómez-Hernández and Wen, 1994). The results show that, the concentration profile is heavily impacted by the connectivity of the conductivity channel structures. For the EnKF, the ensemble mean map of concentration does not show much improvement after head data conditioning. Whereas, for the EnPAT, the reproduction of the main patterns in the reference concentration map is remarkable. This is due to the ability of the EnPAT to replicate the channel structures. Also note the reduction in uncertainty of the predicted concentration profile after conditioning to head data using the EnPAT (see the variance maps).

4. Discussion

The EnPAT was demonstrated as a promising approach to integrate dynamic data into conductivity fields not necessarily multi-Gaussian; however, for practical applications, some issues have to be addressed as discussed next:

- The EnPAT is an extension of the direct sampling MPS method, thus the same issues confronting DS must be addressed in the EnPAT also. For example, the maximum number of conditioning data in a pattern would define the size of pattern; the larger the number of conditioning data, the better the reproduction of patterns from the training images, but also the larger the computational demand of the algorithm. Thus, there is a trade-off between the quality of the model and computational cost. This is also true when specifying distance tolerance values (see Mariethoz et al. (2010b)).

- The EnPAT is a multiple-point statistics-based data integration method. The matched pattern is directly sampled from the ensemble of training images, and thus the curvilinear structures could be

preserved through the process of data conditioning. This distinguishes the method from non-Gaussian particle filter methodology (Moradkhani et al., 2012; Chang et al., 2012) where each particle (i.e., the realization) is sampled from the posterior distribution by evaluating the likelihood. As we know, particle filtering can only be applied to low-dimensional cases because of the problem of filter divergence. Applying particle filters to high-dimensional case is still a very active research topic.

- The exclusive parameter in the EnPAT, which does not appear in DS is the number of pilot points. It was used to reduce computational cost and to improve image quality. Specifically, conductivity and head values are simulated at the pilot point locations at first, and then an MPS method (Direct Sampling here) is used to fill in the spatial patterns in the rest of the domain. The locations of the pilot points are randomly distributed from one realization to the next as well as from one time step to the next one. Li et al. (2013) presented an extensive sensitivity analysis on the choice of the optimum number of pilot points.

- As in the EnKF, the correlation between parameter and state variables are explicitly estimated through the ensemble of training images in EnPAT. Thus, the number of ensemble models is a key parameter that will determine the accuracy of the inferred correlations: the larger the ensemble, the better the estimation, but also the larger the computational cost.

- Measurement error is not considered in the current implementation of the EnPAT algorithm. However, it is straightforward to integrate it into the algorithm. It simply requires the redefinition of the piezometric head distance, such that when the prediction is within a certain interval about the measured value (this interval would be defined by the measurement error), the distance is set equal to zero.

- In the EnKF, multiple parameters can be easily handled by adding additional variables into the joint state vector. For example, Li et al. (2012a) simultaneously updated both porosity and conductivity by conditioning to both piezometric head and concentration data. However, in the EnPAT, this would be very challenging within the current framework. Conceptually, the implementation is quite straightforward by incrementing the variables in the joint pattern, but in practice, a conditional pattern with many variables would be very difficult to match. A hierarchical pattern matching approach could be a possible way to alleviate the strong constraints imposed by a multiple variable pattern match requirement.

- As we know, the EnKF is optimal as a Gaussian-based (two-point statistics) data assimilation methodology, while the EnPAT is a non-Gaussian (multiple-point statistics) method. We also know that the

computational cost of multiple-point statistics is much higher than traditional two-point statistics. Development of approaches (such as those based on cumulants) to reduce the computational cost of multiple-point statistics algorithm is currently a very active research topic. The computational cost of EnPAT for each time step is almost the same as generating the channelized realizations without conditioning on head data using direct sampling MPS method, which itself is much higher than the computational cost used in the two-point covariance based EnKF.

- In the current implementation of the EnPAT, we assume that there is no uncertainty associated with the training image. Specifically, the uncertainties of facies and the corresponding hydraulic conductivity within each facies in the training image are not considered in the comparison. At each time step, the interpolation of simulated values at the pilot points relies on the prior training image. If the training image is wrong, the EnPAT could not identify the structures, accordingly. How to integrate the uncertainty of prior training image would be the future research direction, but the ability to reproduce the curvilinear structures with known training image would be the first step.

5. Conclusions

We compared the EnKF and the EnPAT methods for inverse modeling of a conductivity field characterized by curvilinear channel structures. The EnKF has been widely used in petroleum engineering and hydrogeology over the past decade. The remarkable advantages of the EnKF are the capability to handle multiple parameters, computational efficiency, real-time data assimilation and the ease with which it can be coupled with any forward simulator. However, one significant drawback of the EnKF is that, because it is based on two-point statistics (i.e., covariances) it is optimal only for linear state functions and parameters following multi-Gaussian distributions. As we show in a synthetic subsurface flow and transport example, the EnKF performs poorly when dealing with channelized aquifers with bimodal distribution of conductivity. As an alternative to the EnKF, the EnPAT method was proposed to condition to dynamic data without the limitation that state variables should follow a multi-Gaussian distribution. It is based on a sequential simulation paradigm similar to any MPS algorithm with the exception that there is an ensemble of training images for both conductivities and piezometric heads.

The application of the EnKF and the EnPAT to a synthetic case shows that the EnPAT yields a more accurate characterization of conductivity than the EnKF, and more importantly, the updated models obtained using the EnPAT honor the prior geologic features exhibited in the training image. As a consequence,

the EnPAT has much better accuracy of transport prediction than EnKF. These results stress the capability of EnPAT for conditioning MPS generated conductivity field to observed dynamic data.

For comparison purposes we also implemented a rejection sampling algorithm to generate inverse conditional realizations of conductivity. While the method works comparably to the EnPAT, its cost was prohibitively large in comparison with the EnPAT.

Acknowledgements. The first three authors gratefully acknowledge the financial support by U.S. Department of Energy through project DE-FE0004962. The fourth author acknowledges the financial support by Spanish Ministry of Economy and Competitiveness through project CGL2011-23295. We thank the guest editor Prof. Dr. Harrie-Jan Hendricks Franssen, as well as the reviewer Prof. Alberto Guadagnini and two anonymous reviewers for their comments, which substantially improved the manuscript.

References

Aanonsen, S., Nævdal, G., Oliver, D., Reynolds, A., Valles, B., 2009. The ensemble kalman filter in reservoir engineering–a review. Spe Journal 14 (3), 393–412.

Bailey, R., Baù, D., 2010. Ensemble smoother assimilation of hydraulic head and return flow data to estimate hydraulic conductivity distribution. Water Resources Research 46 (12).

Burgers, G., van Leeuwen, P., Evensen, G., 1998. Analysis scheme in the ensemble Kalman filter. Monthly Weather Review 126, 1719–1724.

Camporese, M., Cassiani, G., Deiana, R., Salandin, P., 2011. Assessment of local hydraulic properties from electrical resistivity tomography monitoring of a three-dimensional synthetic tracer test experiment. Water Resources Research 47 (12).

Chang, S.-Y., Chowhan, T., Latif, S., 2012. State and parameter estimation with an sir particle filter in a three-dimensional groundwater pollutant transport model. Journal of Environmental Engineering 138 (11), 1114–1121.

Chen, Y., Zhang, D., 2006. Data assimilation for transient flow in geologic formations via ensemble Kalman filter. Advances in Water Resources 29 (8), 1107–1122.

Dovera, L., Della Rossa, E., 2011. Multimodal ensemble kalman filtering using gaussian mixture models. Computational Geosciences 15 (2), 307–323.

Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. Journal of Geophysical Research: Oceans (1978–2012) 99 (C5), 10143–10162.

Evensen, G., 2003. The ensemble Kalman filter: Theoretical formulation and practical implementation. Ocean dynamics 53 (4), 343–367.

Gharamti, M., Hoteit, I., 2014. Complex step-based low-rank extended kalman filtering for state-parameter estimation in subsurface transport models. Journal of Hydrology 509, 588–600.

Gharamti, M., Kadoura, A., Valstar, J., Sun, S., Hoteit, I., 2014a. Constraining a compositional flow model with flow-chemical data using an ensemble-based kalman filter. Water Resources Research 50 (3), 2444–2467.

Gharamti, M., Valstar, J., Hoteit, I., 2014b. An adaptive hybrid enkf-oi scheme for efficient state-parameter estimation of reactive contaminant transport models. Advances in Water Resources 71, 1–15.

Gómez-Hernández, J. J., Cassiraga, E. F., 1994. Theory and practice of sequential simulation. In: Armstrong, M., Dowd, P. (Eds.), Geostatistical Simulations. Kluwer Academic Publishers, pp. 111–124.

Gómez-Hernández, J. J., Journel, A. G., 1993. Joint sequential simulation of multigaussian fields. In: Geostatistics Troia °Ø92. Springer, pp. 85–94.

Gómez-Hernández, J. J., Sahuquillo, A., Capilla, J. E., 1997. Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data, 1, Theory. Journal of Hydrology 203 (1–4), 162–174.

Gómez-Hernández, J. J., Wen, X.-H., 1994. Probabilistic assessment of travel times in groundwater modeling. J. of Stochastic Hydrology and Hydraulics 8 (1), 19–56.

Gu, Y., Oliver, D., 2006. The ensemble Kalman filter for continuous updating of reservoir simulation models. Journal of Energy Resources Technology 128, 79.

Guadagnini, A., Guadagnini, L., Tartakovsky, D. M., Winter, C., 2003. Random domain decomposition for flow in heterogeneous stratified aquifers. Stochastic Environmental Research and Risk Assessment 17 (6), 394–407.

Hamill, T., Whitaker, J., Snyder, C., 2001. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. Monthly Weather Review 129, 2776–2790.

Hendricks Franssen, H., Kinzelbach, W., 2008. Real-time groundwater flow modeling with the Ensemble Kalman Filter: Joint estimation of states and parameters and the filter inbreeding problem. Water Resources Research 44 (9), W09408.

Hu, L. Y., Chugunova, T., 2008. Multiple-point geostatistics for modeling subsurface heterogeneity: A comprehensive review. Water Resources Research 44 (11).

Hu, L. Y., Zhao, Y., Liu, Y., Scheepens, C., Bouchard, A., 2013. Updating multipoint simulations using the ensemble kalman filter. Computers & Geosciences 41, 7–15.

Jafarpour, B., Khodabakhshi, M., 2011. A probability conditioning method (PCM) for nonlinear flow data integration into multipoint statistical facies simulation. Mathematical Geosciences 43 (2), 133–164.

Leng, C., Yeh, H., 2003. Aquifer parameter identification using the extended Kalman filter. Water Resources Research 39 (3), 1062.

Li, L., Srinivasan, S., Zhou, H., Gómez-Hernández, J. J., 2013. A pilot point guided pattern matching approach to integrate dynamic data into geological modeling. Advances in Water Resources 62, 125–138.

Li, L., Srinivasan, S., Zhou, H., Gómez-Hernández, J. J., 2014. Simultaneous estimation of geologic and reservoir state variables within an ensemble-based multiple-point statistic framework. Mathematical Geosciences 46 (5), 597–623.

Li, L., Zhou, H., Gómez-Hernández, J. J., Hendricks Franssen, H.-J., 2012a. Jointly mapping hydraulic conductivity and porosity by assimilating concentration data via ensemble kalman filter. Journal of Hydrology 428, 152–169.

Li, L., Zhou, H., Hendricks Franssen, H., Gómez-Hernández, J. J., 2011. Modeling transient flow by coupling ensemble kalman filtering and upscaling. Water Resources Research, doi:10.1029/2010WR010214.

Li, L., Zhou, H., Hendricks Franssen, H. J., Gómez-Hernández, J. J., 2012b. Groundwater flow inverse modeling in non-multigaussian media: performance assessment of the normal-score ensemble kalman filter. Hydrology and Earth System Sciences 16 (2), 573–590.

Mariethoz, G., Renard, P., Caers, J., 2010a. Bayesian inverse problem and optimization with iterative spatial resampling. Water Resources Research 46 (11).

Mariethoz, G., Renard, P., Straubhaar, J., 2010b. The direct sampling method to perform multiple-point geostatistical simulaitons. Water Resources Research 46.

Meerschman, E., Pirot, G., Mariethoz, G., Straubhaar, J., Meirvenne, M., Renard, P., 2012. A practical guide to performing multiple-point statistical simulations with the direct sampling algorithm. Computers & Geosciences.

Moradkhani, H., DeChant, C. M., Sorooshian, S., 2012. Evolution of ensemble data assimilation for uncertainty quantification using the particle filter-markov chain monte carlo method. Water Resources Research 48 (12).

Panzeri, M., Della Rossa, E., Dovera, L., Riva, M., Guadagnini, A., 2014a. Integration of markov mesh models and ensemble data assimilation in reservoirs with complex geology. In: ECMOR XIV-14th European conference on the mathematics of oil recovery.

Panzeri, M., Riva, M., Guadagnini, A., Neuman, S., 2013. Data assimilation and parameter estimation via ensemble kalman filter coupled with stochastic moment equations of transient groundwater flow. Water Resources Research 49 (3), 1334–1344.

Panzeri, M., Riva, M., Guadagnini, A., Neuman, S., 2014b. Comparison of ensemble kalman filter groundwater-data assimilation methods based on stochastic moment equations and monte carlo simulation. Advances in Water Resources 66, 8–18.

Panzeri, M., Riva, M., Guadagnini, A., Neuman, S., 2014c. Enkf coupled with groundwater flow moment equations applied to lauswiesen aquifer, germany. Journal of Hydrology.

Ping, J., Zhang, D., 2013. History matching of fracture distributions by ensemble kalman filter combined with vector based level set parameterization. Journal of Petroleum Science and Engineering 108, 288–303.

Riva, M., Guadagnini, A., Fernandez-Garcia, D., Sanchez-Vila, X., Ptak, T., 2008. Relative importance of geostatistical and transport models in describing heavily tailed breakthrough curves at the lauswiesen site. Journal of contaminant hydrology 101 (1), 1–13.

Sarma, P., Chen, W., 2009. Generalization of the ensemble kalman filter using kernels for nongaussian random fields. In: SPE Reservoir Simulation Symposium.

Satija, A., Caers, J., 2015. Direct forecasting of subsurface flow response from non-linear dynamic data by linear least-squares in canonical functional principal component space. Advances in Water Resources.

Scheidt, C., Renard, P., Caers, J., 2014. Prediction-focused subsurface modeling: Investigating the need for accuracy in flow-based inverse modeling. Mathematical Geosciences, 1–19.

Schöniger, A., Nowak, W., Hendricks Franssen, H. J., 2012. Parameter estimation by ensemble Kalman filters with transformed data: approach and application to hydraulic tomography. Water Resources Research 48 (4).

Strebelle, S., 2002. Conditional simulation of complex geological structures using multiple-point statistics. Mathematical Geology 34 (1), 1–21.

Sun, A. Y., Morris, A. P., Mohanty, S., Jul. 2009. Sequential updating of multimodal hydrogeologic parameter fields using localization and clustering techniques. Water Resources Research 45, 15 PP.

Wen, X., Capilla, J., Deutsch, C., Gómez-Hernández, J., Cullick, A., 1999. A program to create permeability fields that honor single-phase flow rate and pressure data. Computers & Geosciences 25 (3), 217–230.

Wen, X., Chen, W., 2006. Real-time reservoir model updating using ensemble kalman filter with confirming option. SPE Journal 11 (4), 431–442.

Wen, X.-H., Tran, T. T., Behrens, R. A., Gómez-Hernández, J. J., 2000. Production data integration in sand/shale reservoirs using sequential self-calibration and geomorphing: A comparison. Spe 63063 -.

Winter, C., Tartakovsky, D., Guadagnini, A., 2003. Moment differential equations for flow in highly heterogeneous porous media. Surveys in Geophysics 24 (1), 81–106.

Winter, C., Tartakovsky, D. M., Guadagnini, A., 2002. Numerical solutions of moment equations for flow in heterogeneous composite aquifers. Water resources research 38 (5), 13–1.

Zhou, H., Gómez-Hernández, J., Hendricks Franssen, H., Li, L., 2011. An approach to handling non-gaussianity of parameters and state variables in ensemble kalman filtering. Advances in Water Resources 34 (7), 844–864.

Zhou, H., Gómez-Hernández, J., Li, L., 2012. A pattern-search-based inverse method. Water Resources Research 48 (3), W03505.

Zhou, H., Gómez-Hernández, J. J., Li, L., 2014. Inverse methods in hydrogeology: evolution and recent trends. Advances in Water Resources 63, 22–37.

Table 1: Aquifer configuration and parameters used in the DS and the EnPAT.

Aquifer size	$50 \times 50 \times 1\ m$
Gridblock size	$1 \times 1 \times 1\ m$
Log-conductivity mean	-0.12 md^{-1}
Log-conductivity standard deviation	2.51 md^{-1}
Porosity	0.3
Specific storage	0.02 m^{-1}
Number of time steps	10
Total simulation time	30 days
Number of observation k	9
Number of observation wells	9
Initial head	0 m
Boundary conditions	no flow (north and south); $h = 0$ (east and west)
Injection well	$25m^3/d$
Number of realizations	500
Search radius for k (EnPAT & DS)	25 m
Search radius for h (EnPAT)	25 m
Max. number of element in the pattern for k (EnPAT & DS)	15
Max. number of element in the pattern for h (EnPAT)	15
Distance function for k (EnPAT & DS)	Weighted Euclidian
Distance function for h (EnPAT)	Weighted Euclidian
Distance tolerance for k (EnPAT & DS)	0.05
Distance tolerance for h (EnPAT)	0.0
Fraction of scan on training image (DS)	0.5
Number of pilot points (EnPAT)	300

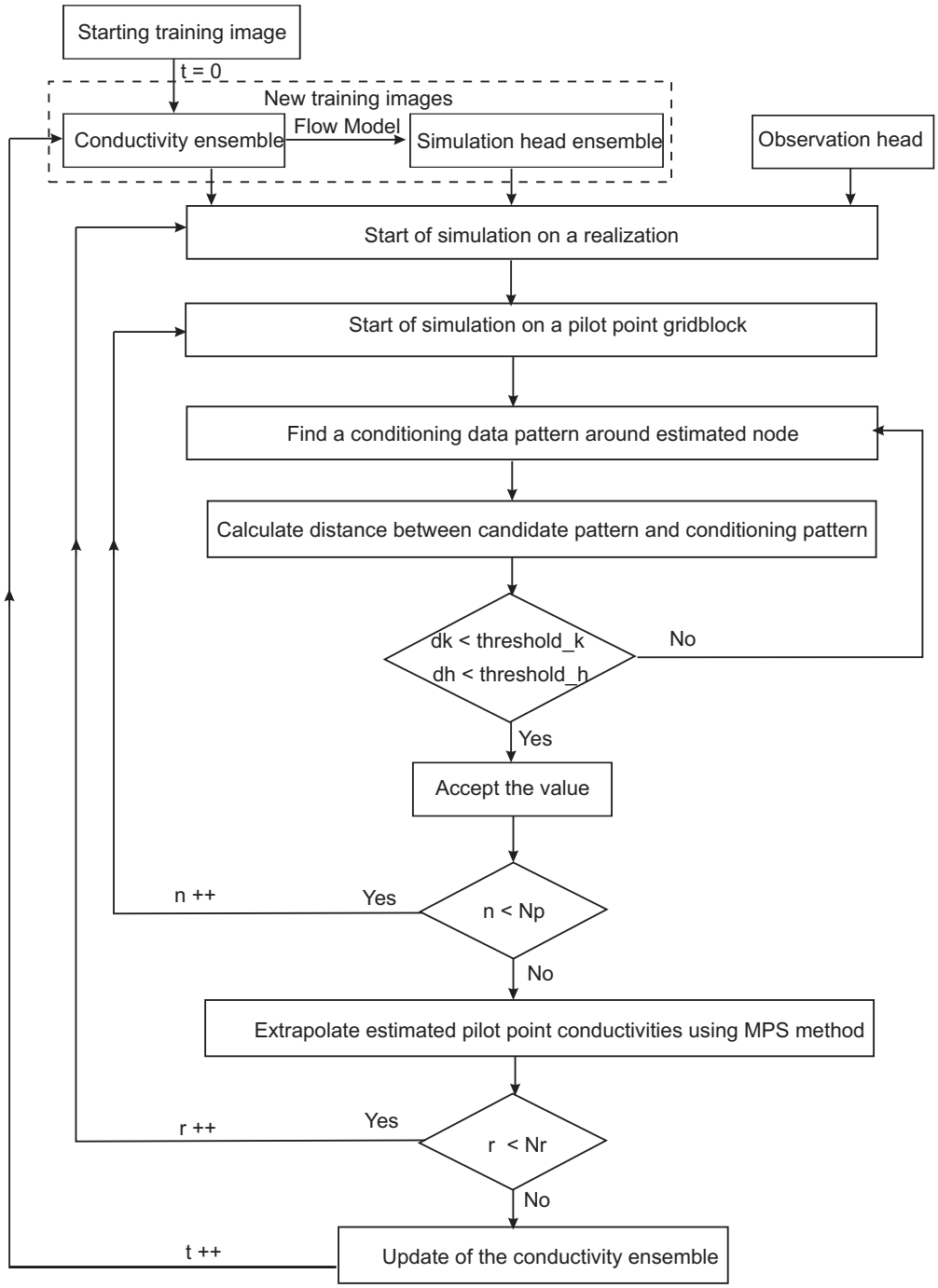

Figure 1: Flowchart of the EnPAT algorithm.

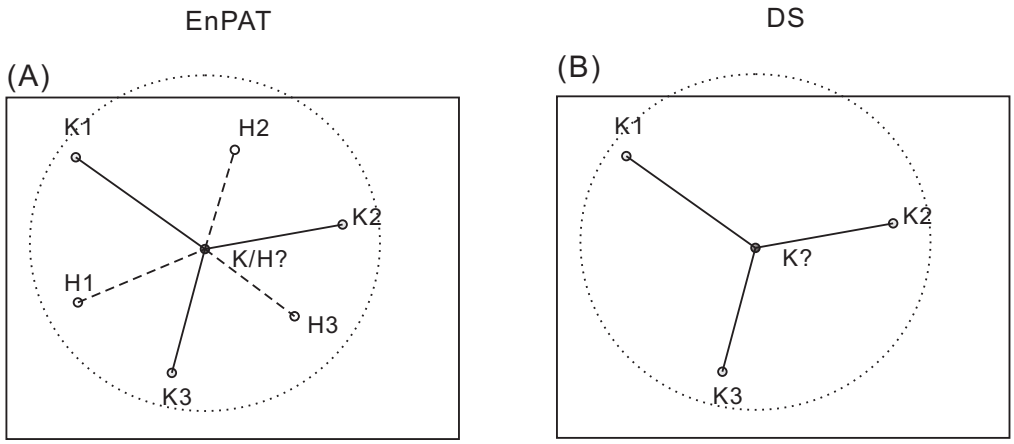

Figure 2: Definition of pattern: (A) ensemble pattern matching, and (B) direct sampling.

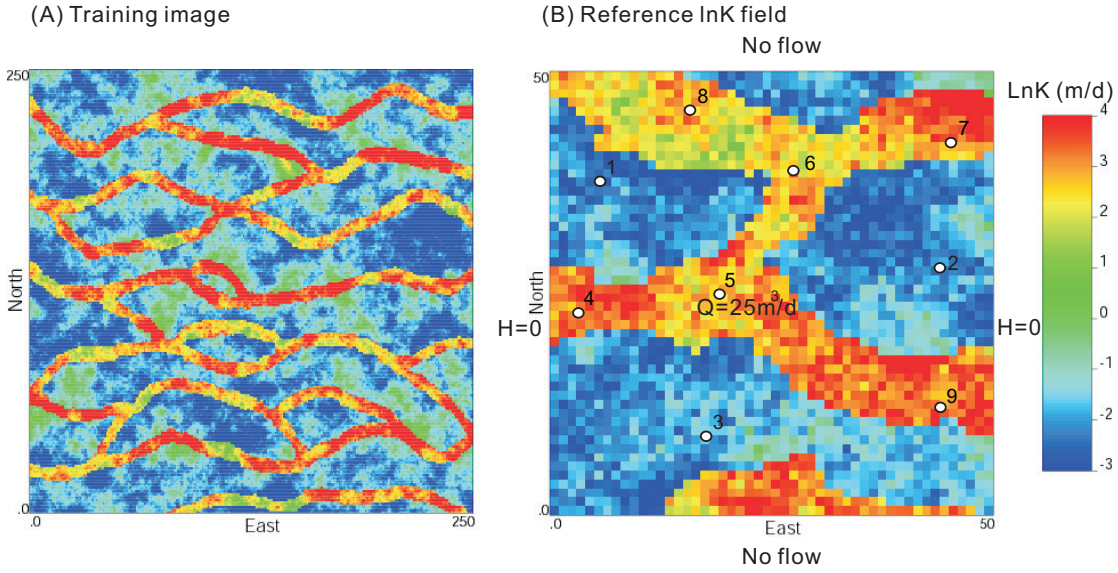

Figure 3: (A) Training image (B) Reference conductivity, boundary conditions of flow model and observation wells.

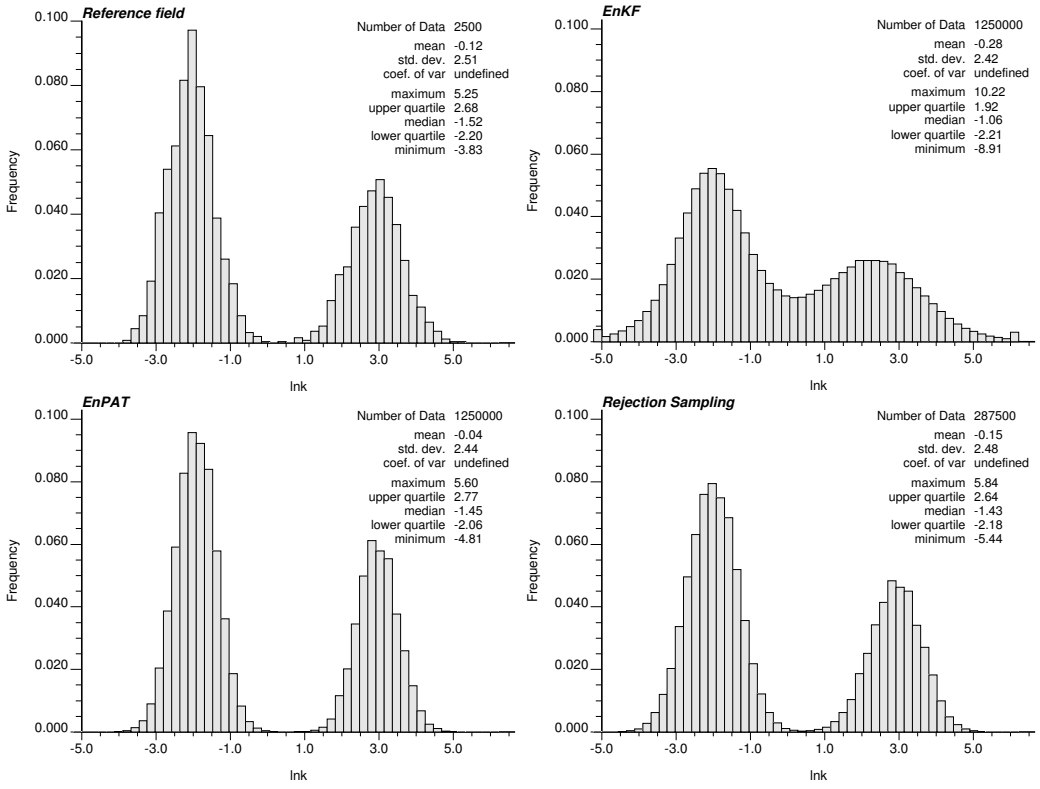

Figure 4: Log-conductivity histograms for the reference conductivity and for the ensemble of conductivities obtained by the EnKF, the EnPAT and rejection sampling after data conditioning.

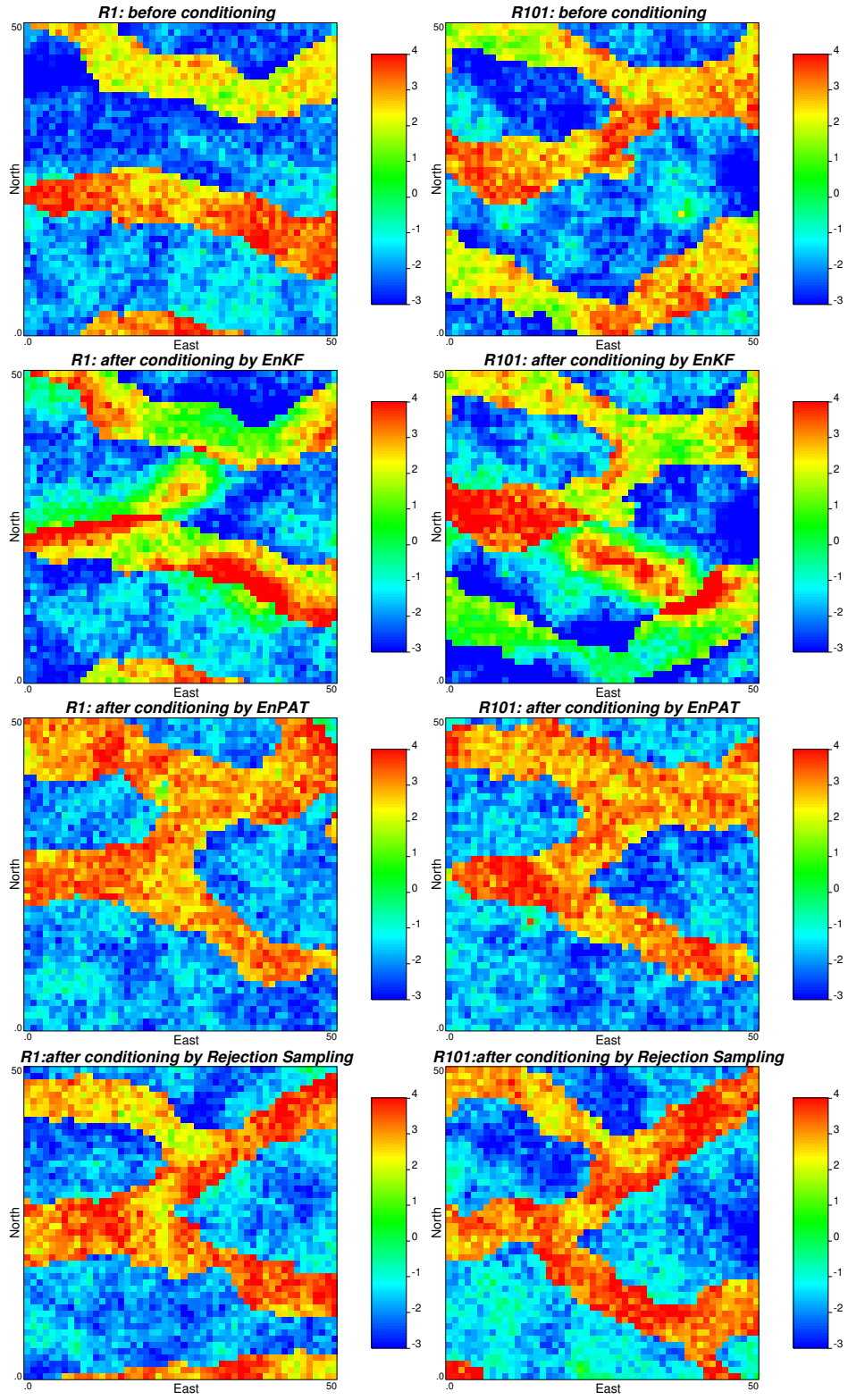

Figure 5: Randomly selected two individual realizations before and after head data conditioning using the EnKF, the EnPAT and rejection sampling.

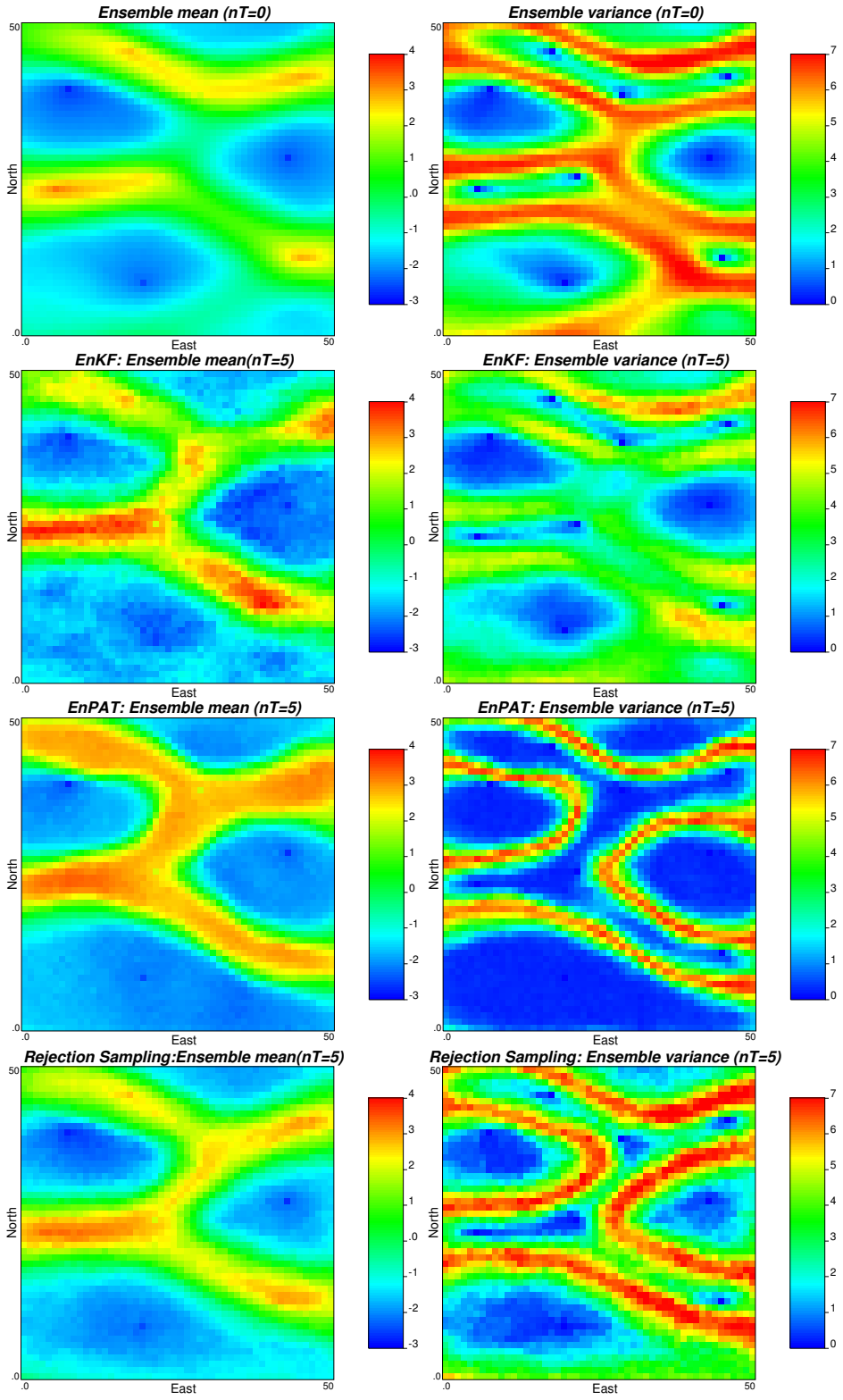

Figure 6: Ensemble mean and variance of log-conductivity before and after head data conditioning using the EnKF, the EnPAT and rejection sampling.

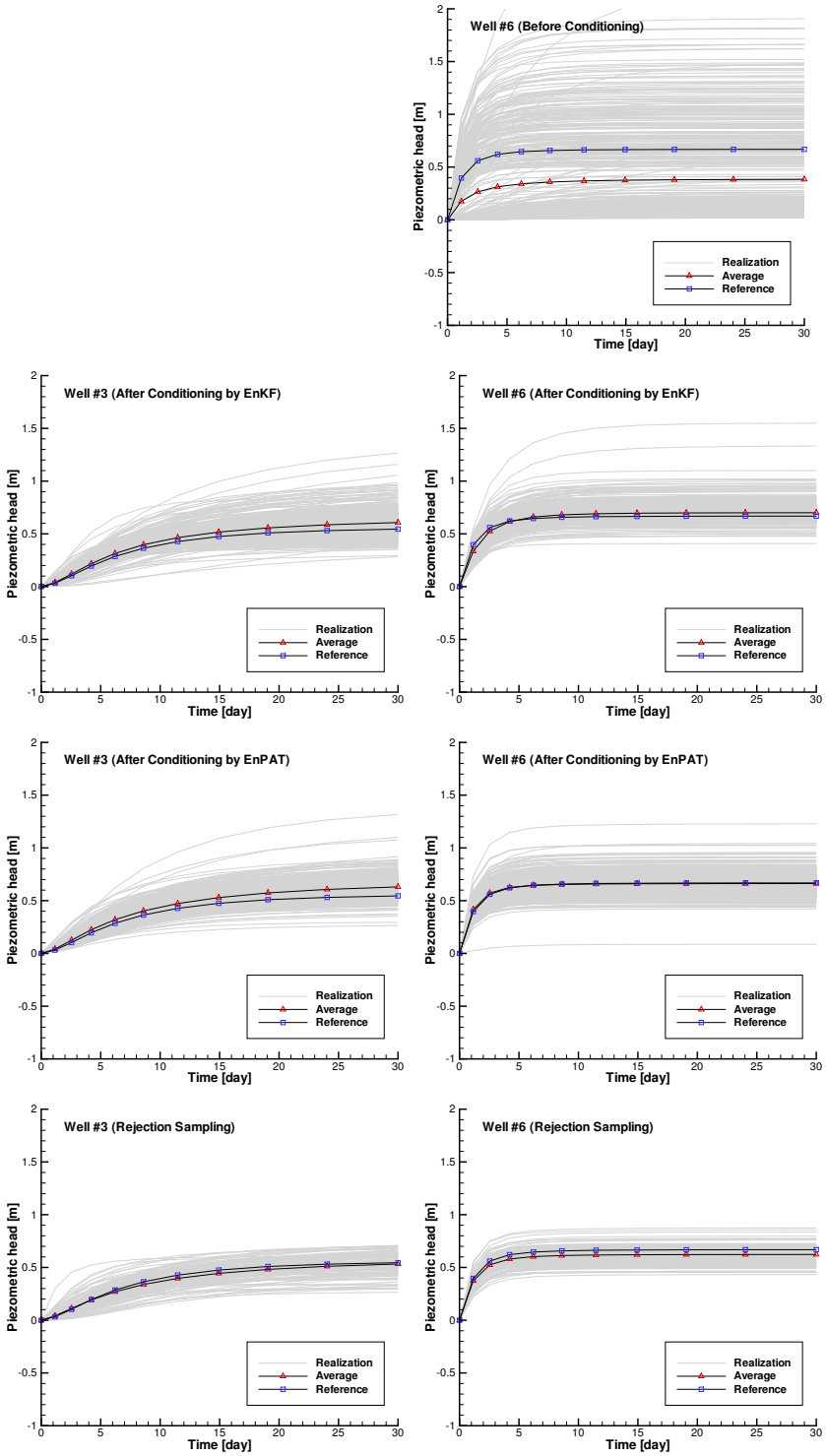

Figure 7: The simulated head at two wells before and after head data conditioning using the EnKF, the EnPAT and rejection sampling.

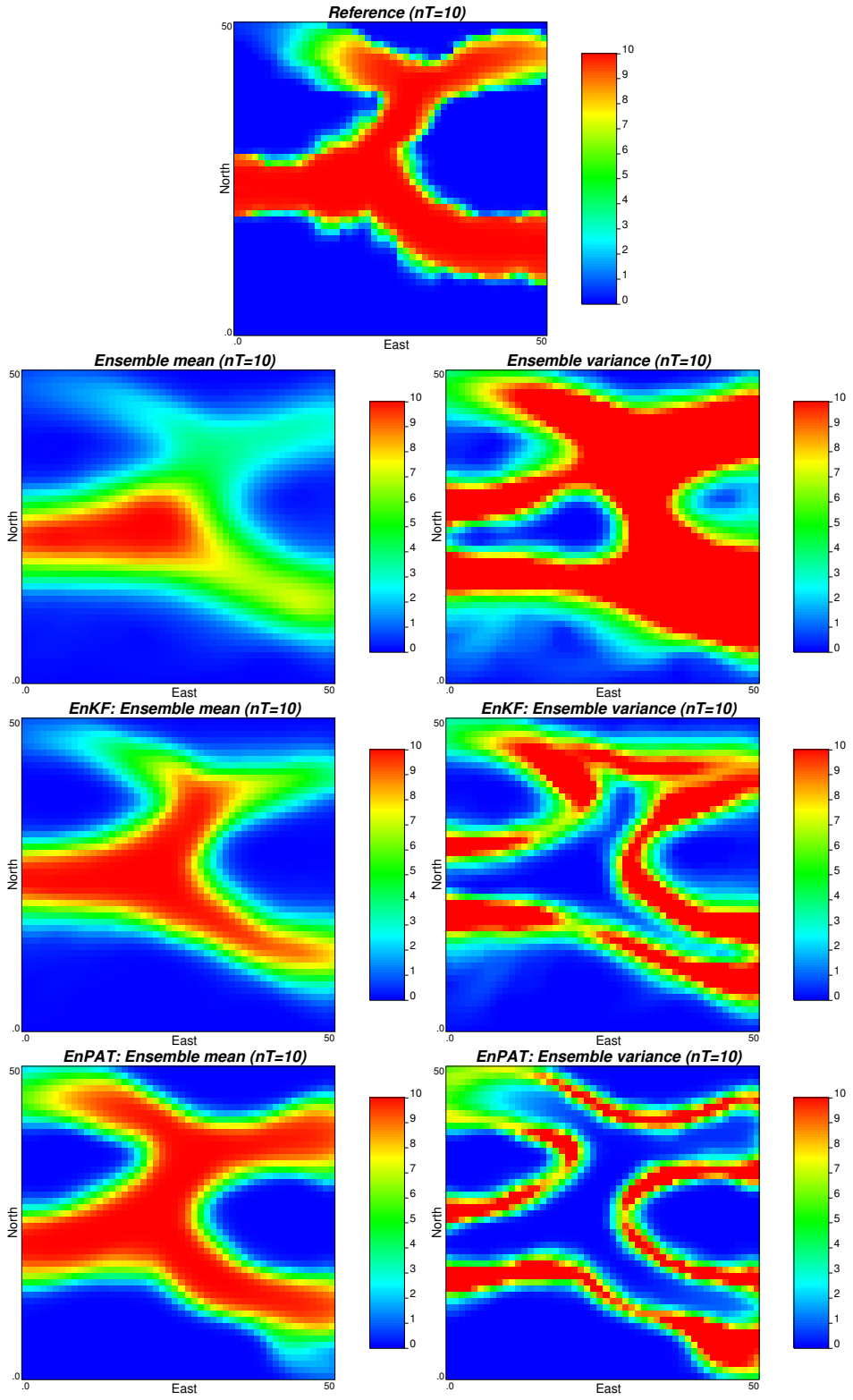

Figure 8: Ensemble mean and variance of concentration ($nT = 10$) before and after head data conditioning using the EnKF and the EnPAT.