

Document downloaded from:

<http://hdl.handle.net/10251/52671>

This paper must be cited as:

Li, L.; Srinivasan, S.; Zhou, H.; Gómez-Hernández, JJ. (2014). Simultaneous estimation of both geologic and reservoir state variables within an ensemble-based multiple-point statistic framework. *Mathematical Geosciences*. 46(5):597-623. doi:10.1007/s11004-013-9504-z.



The final publication is available at

<http://dx.doi.org/10.1007/s11004-013-9504-z>

Copyright Springer Verlag (Germany)

# Simultaneous estimation of both geologic and reservoir state variables within an ensemble-based multiple-point statistic framework

Liangping Li · Sanjay Srinivasan · Haiyan Zhou · J. Jaime Gómez-Hernández

Received: date / Accepted: date

**Abstract** Assessment of uncertainty due to inadequate data and imperfect geological knowledge is an essential aspect of the subsurface model building process. In this work, a novel methodology for characterizing complex geological structures is presented that integrates dynamic data. The procedure results in the assessment of uncertainty associated with the predictions of flow and transport. The methodology is an extension of a previously developed pattern search-based inverse method that models the spatial variation in flow parameters by searching for patterns in an ensemble of reservoir models. More specifically, the pattern-searching algorithm is extended in two directions: (1) state values (such as piezometric head) and parameters (such as conductivities) are simultaneously and sequentially estimated, which implies that real-time assimilation of dynamic data is possible as in ensemble filtering approaches; (2) both the estimated parameter and state variables are considered when pattern searching is implemented. The new scheme results in two main advantages - better characterization of parameters, especially for delineating small scale features, and an ensemble of head states that can be used to update the parameter field using the dynamic data at the next instant, without running expensive flow simulations. An efficient algorithm for pattern search is developed,

---

Liangping Li  
Center for Petroleum and Geosystems Engineering Research, University of Texas at Austin, 78712, Austin, USA  
E-mail: liangpingli@utexas.edu

Sanjay Srinivasan  
Center for Petroleum and Geosystems Engineering Research, University of Texas at Austin, 78712, Austin, USA  
E-mail: sanjay.srinivasan@enr.utexas.edu

Haiyan Zhou  
Center for Petroleum and Geosystems Engineering Research, University of Texas at Austin, 78712, Austin, USA  
E-mail: haiyanzhou@utexas.edu

J. Jaime Gómez-Hernández  
Research Institute of Water and Environmental Engineering, Universitat Politècnica de València, 46022, Valencia, Spain  
E-mail: jgomez@upv.es

which works with a flexible search radius and it can be optimized for the estimation of either large- or small-scale structures. Synthetic examples are employed to demonstrate the effectiveness and robustness of the proposed approach.

**Keywords** multiple-point statistics · pattern · inverse method · ensemble-based method · history matching

## 1 Introduction

In the past several decades, inverse methods have been developed and increasingly utilized to estimate geological parameters and/or state values (such as conductivity, porosity and piezometric head) by conditioning to the observed state data (such as head and concentration). The goal of these methods is to improve the predictions of flow and transport. Examples are the pilot point method (Ramarao et al 1995), the sequential self-calibration (Gómez-Hernández et al 1997), and the ensemble Kalman filter (EnKF) (Evensen 2003). All the approaches mentioned above are optimal if the prior parameter and/or state can be modeled by a multi-Gaussian distribution meaning that their spatial variability is fully described by a two-point covariance. However, reproduction of complex curvilinear structures is not possible using only a two-point covariance. Examples of reservoirs exhibiting such structures include fluvial carbonate and fractured systems (Huang and Srinivasan 2012).

An alternative to covariance-constrained methods is the multiple-point statistics (MPS) method, which is able to characterize curvilinear structures. In this approach, the local conditional probability is derived from a training image, and the modeled reservoir is sampled from the conditional probability. The training image as the conceptual model of the geological structure could be based on a rock outcrop or built from expert knowledge on the reservoir (Guardiano and Srivastava 1993). The first effective MPS algorithm, SNESIM, was developed by Strebelle (2002). Other variants of the MPS algorithm are available in the literature, for example, FILTERSIM (Zhang et al 2006), SIMPAT (Arpat 2005), DS (Mariethoz et al 2010b), IMPALA (Straubhaar et al 2011), CCSIM (Tahmasebi et al 2012a), GROWTHSIM (Huang and Srinivasan 2012), MPCCA (Abdollahifard and Faez 2013). An extensive review of MPS method can be found in the paper by Hu and Chuginova (2008).

History matching with MPS calls for an effective inverse method that can not only honor the observed state data, but also preserve the curvilinear geological structure characterizing the reservoir. Caers (2003) developed the probability perturbation technique, in which the probability of conductivity value at a location conditioned to the dynamic data is modeled using a deformation parameter. This conditional probability is subsequently merged with the pattern inferred from the training image (Caers 2002; Hoffman and Caers 2005; Kashib and Srinivasan 2006). Alcolea and Renard (2010) developed a block moving window algorithm, an extension of the block Markov chain Monte Carlo method by Fu and Gómez-Hernández (2009), for conditioning MPS simulations to piezometric head data as well as to connectivity data. The posterior distribution of parameter is obtained by sampling the prior models, which have already been conditioned to the observed static data within each block. Mariethoz et al (2010a) proposed the iterative spatial

resampling method to preserve the observation data, but this method is computationally expensive. Jafarpour and Khodabakhshi (2011) developed the probability conditioning method, in which the dynamic data are first conditioned by EnKF to derive the parameter mean values, and then used these as soft data to regenerate the parameter realizations using MPS. Zhou et al (2012) developed a pattern searching inverse method to estimate both static and state parameters within an MPS framework. They extend the Direct Sampling (DS) method by Mariethoz et al (2010b) for inverse modeling; for this purpose they work with an ensemble of realizations of conductivity and their associated heads as multiple training images, this allows them to take into account patterns in the spatial fields of measured heads and conductivities when sampling with the DS algorithm.

The use of highly instrumented wells for both groundwater and hydrocarbon recovery has become common. Dynamic data continuously obtained from on-line sensors along wells require an inverse model that can integrate these data in real-time so that the model is constantly updated to reflect the available data. Traditional inverse methods generally require the model to be rerun from the beginning once new observed data are available, resulting in high computational cost and therefore are impractical for real-time data assimilation. On the other hand, if the state variables, such as piezometric head are updated simultaneously with the reservoir parameters, data assimilation would be possible by simply re-running the models from the previously updated step. This is done in the EnKF method, which is able to simultaneously estimate the parameter and state variables using incremental observed dynamic data and thus renders possible real-time data assimilation. The EnKF is gaining popularity in various disciplines such as meteorology, petroleum engineering and hydrogeology (Hamill et al 2001; Wen and Chen 2006; Gu and Oliver 2006; Chen and Zhang 2006; Hendricks Franssen and Kinzelbach 2008; Li et al 2012). However, EnKF has an optimal solution only if the parameter follows a multiGaussian distribution and if the state transfer function is linear (Aanonsen et al 2009). Sun et al (2009) coupled the EnKF and Gaussian mixture models to handle non-Gaussian parameter distributions. Zhou et al (2011) proposed to transform the non-multiGaussian parameter and state to Gaussian ones through a normal-score transformation approach and then perform updating of the transformed variables at each time step. However, their procedure utilizes a univariate Gaussian transformation. MultiGaussianity is not ensured by this procedure and the method may yield sub-optimal solutions. Hu et al (2012) proposed to update the uniform random field, which is used to draw outcomes from the conditional distribution in the MPS algorithm, using the EnKF.

In this work, a novel inverse method: Ensemble PATtern searching method (EnPAT) is proposed, which is an extension of the pattern searching inverse method previously proposed by Zhou et al (2012). Unlike the previous work focusing only on the estimation of geologic parameter such as conductivity, in this study, both the geologic parameter and reservoir state variables such as piezometric head, at each conditioning time step, are explicitly estimated through a pattern searching scheme. The purpose of simultaneous estimation of parameter and state is twofold: to make updating of parameter field using incremental dynamic data possible by executing the flow simulator only over the incremental time step, mimicking the ensemble filtering approaches, and to better preserve the relationship between the geologic and state variables over multiple length scales. In order to accomplish this, a flexible pattern search strategy is implemented without specifying template

as in the traditional MPS method. Moreover, unlike the previous study (Zhou et al 2012) that only works on a categorical parameter field, here the proposed method is demonstrated both in the cases of categorical and continuous parameter fields.

The remaining of the paper is organized as follows. First, Sect. 2 introduces the proposed pattern searching inverse methodology. Next, in Sect. 3, synthetic examples are used to demonstrate the accuracy and effectiveness of the method. In Sect. 4, there is a general discussion of the results. Finally, the main results and conclusions of this paper are summarized.

## 2 Methodology

EnPAT is fully based on multiple-point pattern search and reproduction. The algorithm utilizes DS for modeling curvilinear and complex geologic structures and for representing spatial variations in state variables conditioned to available dynamic data. For understanding EnPAT better, the DS algorithm (Mariethoz et al 2010b) will first briefly revisited, and then the pattern searching inverse method EnPAT will be introduced.

### 2.1 Direct Sampling Method

Unlike the MPS algorithm introduced by Strebelle (2002), DS directly samples the patterns (data events) from the training image looking for a match to the conditioning pattern, thereby avoiding the storage of data events prior to simulation. Mariethoz et al (2010b) demonstrated that the proposed method works for both the categorical and continuous variables as well as when multiple attributes are jointly simulated. The basic algorithm of DS is as follows:

1. Obtain a training image.
2. Start the loop for the simulation of the  $k$ th realization.
3. Define a random path:  $P_i(x)$ ,  $i = 1, 2, \dots, n$   $x \in D$ , with  $D$  being the reservoir grid that has been discretized into  $n$  grid blocks.
4. To simulate the  $i$ th parameter value:  $Z_i(x)$ , perform the following steps:
  - (a) If the parameter value at the  $i$ th location is known by measurement (for example, the location of hard data), go to step 5.
  - (b) Build the conditioning pattern within a given maximum search radius; the number of data retained in the pattern is less than the predefined maximum number of conditioning data. Note that the conditioning pattern includes both measured data and previously simulated values (Fig. 1A and B).
  - (c) Define a random search path on the training image:  $P_j(y)$ ,  $j = 1, 2, \dots, m$  with  $y \in TI$ ,  $TI$  being the training image grid that has been discretized into  $m$  grid blocks.
  - (d) Loop on  $j$  in order to find the matched candidate pattern in the training image,
    - i. Calculate the distance  $d_j$  between the conditioning pattern and the  $j$ th pattern in the training image.
    - ii. If  $d_j$  is smaller than a predefined distance tolerance value, then,  $Z_i(x) = Z_j(y)$ . Go to step 5.

- iii. Otherwise, set  $j = j + 1$ , and go to step i.
- (e) End the loop. If at the end of step (d), a training pattern matching within the tolerance is not found, then choose the pattern with the smallest distance  $d_{min}$ , and assign  $Z_i(x) = Z_{d_{min}}(y)$ . Like other sequential simulations, the simulated  $i$ th parameter value will be used as conditioning data from now on.
- 5. Set  $i = i + 1$  and go to the next grid block along the random path and then step 4 until all block are visited.
- 6. Set  $k = k + 1$  and go to step 2 until all realizations are simulated.

Note that the quality of the simulation results is dependent on a set of user-defined parameters (maximum search radius, maximum number of conditioning data in the pattern, distance function and distance tolerance). An extensive sensitivity analysis of these parameters on the simulation results can be found in the work by Meerschman et al (2013). The size of the pattern is determined by the maximum search radius and maximum number of conditioning data. As described by Mariethoz et al (2010b), at the early stage of the simulation, the conditioning data are usually very sparse resulting in patterns extending over a large area, which is beneficial for the characterization of large scale structures such as channels (Fig. 1A). At the later stage of the simulation, the conditioning data become much more dense and then the size of the conditioning pattern becomes small, helping to reproduce structures on a smaller scale, such as the conductivity within the channel facies or matrix or the shale drapes separating stacked channels (Fig. 1B). In this way, a flexible search strategy is achieved through two parameters: the maximum search radius and maximum number of conditioning data. The distance function and tolerance will be discussed in detail in Sect. 2.3.

## 2.2 Pattern Searching Inverse Method

In many cases, besides the hard conditioning data, auxiliary information such as piezometric head might also be available to characterize the reservoir. Based on idea of the DS, Zhou et al (2012) proposed to consider the pattern including both the static variables (conductivity) and observed state data (piezometric head). In addition, an ensemble of training images and their corresponding state realizations are considered for pattern searching. However, in their approach, only the spatial conductivity field is simulated. As mentioned previously, in order for real-time model updating to be performed efficiently, both the conductivity field as well as the state variables such as head have to be updated simultaneously. This will permit incremental updating to be possible (as in EnKF). Therefore, the pattern searching inverse method is improved by simultaneously simulating both conductivity and state variables such as head using the extended DS algorithm.

Figure 2 shows the flowchart of the proposed algorithm, which can be summarized as follows:

1. Obtain an ensemble of  $N_r$  parameter training images. The total number of assimilation time steps is  $nT$  (i.e., the dynamic data are available at  $nT$  time steps).
2. Start the loop with  $nT = 1$ .

3. Run flow simulations from  $nT - 1$  to  $nT$  to get the corresponding  $N_r$  state realizations. Note that the ensemble parameter at time  $nT - 1$  and state at time  $nT$  will be used as training images in the subsequent pattern search scheme.
4. Start the loop with realization  $k = 1$ .
5. Define a random path:  $P_i(x)$ ,  $i = 1, 2, \dots, n$ ,  $x \in D$  visiting all  $n$  nodes of the reservoir grid.
6. To simulate the reservoir parameter  $Z_i(x)$  and the state variable  $Y_i(x)$  values, start the loop for node  $i = 1$ :
  - (a) If both reservoir parameter and state variable at the  $i$ th node are known by measurement, go to step 6.
  - (b) Build the conditioning pattern within the maximum search radius; the number of data in the pattern should be less than the predefined maximum number of conditioning data. The pattern may be composed of both parameter and state variables (Fig. 1E and F). A different maximum number of conditioning data can be specified for parameter and state if needed. Here, a spiral search scheme centered on the simulation location  $i$  is considered to find the conditioning data within the maximum search radius, in order to save computational cost.
  - (c) Define a random permutation  $P_j$ ,  $j = 1, 2, \dots, N_r$  to visit the  $N_r$  realizations. Look for the matching pattern as follows:
  - (d) Start a loop with realization  $j = 1$ .
    - i. In realization  $P_j$ , retrieve the data matching the conditioning pattern geometry at exactly the same node  $i$ . Note that the training images are visited in a random order but the conditioning pattern location is fixed around the same node (Fig. 3).
    - ii. Calculate the distances  $d_j^X$  for the parameters and  $d_j^Y$  for the state variable between the conditioning pattern and the pattern found in the  $j$ th ensemble realization and its corresponding flow simulation result.
    - iii. If  $d_j^X$  and  $d_j^Y$  are both smaller than the predefined distance tolerance values, then  $Z_i(x) = Z_j(x)$  and  $Y_i(x) = Y_j(x)$ . Go to step 6.
    - iv. Otherwise, set  $j = j + 1$ , and go to step i.
  - (e) Keep track of the realization having the pattern with the smallest distance both  $d_{min}^X$  and  $d_{min}^Y$ . Assign  $Z_i(x) = Z_{d_{min}}(x)$  and  $Y_i(x) = Y_{d_{min}}(x)$ . Like other sequential simulations, the simulated  $i$ th parameter and state values will be used as conditioning data from now on.
7. Set  $i = i + 1$  and go to step 5 until all nodes are visited in the simulation grid.
8. Set  $k = k + 1$  and go to step 3 until all realizations are simulated.
9. Set  $nT = nT + 1$  and go to step 2 until all time steps are simulated.

Note that, for the sake of simplicity, here it is assumed that both parameter and state are measured at the same locations. In general, the hard and soft/auxiliary information need not be available at the same location. For example, in the case with only state available (i.e., without hard data), the conditioning pattern is composed of only state variable at the start of simulation, and then the parameter and state variables both are simulated on the basis of conditioning state pattern. Once the previous estimated parameter is found within the maximum search radius in the subsequent simulations, the conditioning pattern is enlarged to include both the parameter and state variables. This variation of EnPAT algorithm was implemented in the paper by Li et al (2013).

EnPAT shares some features with the traditional implementation of DS: (1) Both methods are based on sequential simulation; (2) both methods are based on multiple-point patterns, ensuring reproduction of curvilinear geologic features as observed in the training image; (3) patterns are directly sampled rather than building a conditional probability distribution and then drawing from it; (4) a flexible search strategy, resulting in a good characterization of large and small scale features (Fig. 1); (5) the quality of pattern reproduction in both DS and EnPAT is controlled by user-defined parameters such as maximum search radius, maximum number of conditioning data in the pattern, distance function and distance tolerance.

However, EnPAT has some exclusive features: (1) by using multiple training images, the patterns are local, that is, a pattern match is searched through the ensemble of training images at the same exact location; this feature allows handling the non-stationarity dependence of the piezometric head on the conductivity, due to the presence of wells, sinks or boundaries (Fig. 3); (2) the pattern includes both parameter values and state variables, thus allowing the inverse modeling of conductivities; (3) because the conditioning pattern is composed of both conductivities and heads, two distances have to be computed, one for the conductivities in the pattern, and the other one for the heads, each one with its own definition and tolerance as will be explained later.

EnPAT is an extension of the work by Zhou et al (2012) resulting in an improved pattern search inverse method. In the algorithm by Zhou et al (2012), the conditioning pattern is composed of both the pattern of conductivity (original data + previously simulated values) as well as the pattern of measured state variable values. Unlike in EnPAT, the size of the pattern that is matched during simulation remains large because it is controlled by the spacing at which the measured state variable values are available. At later stages of simulation, the matching pattern consists of conductivity values in the close vicinity of the simulation node while the measured state variables might still be available at large distances from the simulation node (Fig 1D). Therefore, the small scale correlation between parameter and state can not be captured by this approach. This can be used to explain why the simulated conductivity fields are noisy after conditioning to the state data in the synthetic example of Zhou et al (2012). It is our assertion that combined simulation of both state variables and conductivity patterns will help reduce the noise in the simulated models.

Pattern-based updating of the head state might lead to inconsistencies between the parameter variables and the simulated head state in the form of material balance error. These problems are common to all ensemble updating methods such as EnKF and can be somewhat alleviated by using a larger ensemble in order to learn the relationship between the parameter and the state variables. An alternative is also to periodically perform full physics flow simulations over the entire duration of the data in order to obtain the correct spatial distribution of state variables corresponding to the current state of the parameters. In that case the pattern search algorithm can be viewed as a piecewise interpolation scheme between major updates of the state variables.



### 2.3 Distance Function

The distance function plays a crucial role in both the DS and EnPAT algorithms. It is used to quantify the dissimilarity between the conditional pattern and the candidate patterns found in the training image. As described in Mariethoz et al (2010b) and Zhou et al (2012), two kinds of distance functions are commonly used:

1. Manhattan distance:

- Categorical variables

$$d\{d(x_n), p(x_n)\} = \frac{1}{n} \sum_{i=1}^n a_i \quad d \in [0, 1],$$

$$a_i = \begin{cases} 0, & \text{if } d(x_i) = p(x_i) \\ 1, & \text{otherwise} \end{cases} \quad (1)$$

where  $n$  is the numbers of nodes in the pattern,  $d(x_n)$  is the conditioning pattern,  $p(x_n)$  is the matching pattern,  $x_i$  represents the members of the pattern, either parameter or state variable values.

- Continuous variables

$$d\{d(x_n), p(x_n)\} = \frac{1}{n} \sum_{i=1}^n \frac{|d(x_i) - p(x_i)|}{d_{max}} \quad d \in [0, 1], \quad (2)$$

where  $d_{max}$  is the maximum absolute difference  $|d(x_i) - p(x_i)|$ .

2. Weighted Euclidean distance:

- Categorical variables

$$d\{d(x_n), p(x_n)\} = \frac{1}{\sum_{i=1}^n h_i^{-1}} \sum_{i=1}^n h_i^{-1} a_i \quad d \in [0, 1], \quad (3)$$

where  $h_i$  is the Euclidean distance between node  $i$  and the node being simulated, therefore giving more weight to dissimilarities of the closest nodes.

- Continuous variables

$$d\{d(x_n), p(x_n)\} = \left( \frac{1}{\sum_{i=1}^n h_i^{-1}} \sum_{i=1}^n h_i^{-1} \frac{|d(x_i) - p(x_i)|^2}{d_{max}^2} \right)^{1/2} \quad d \in [0, 1]. \quad (4)$$

From the previous definitions, note that: (1) distance values range from 0 to 1 making it easier to compare patterns and apply tolerances; (2) the choice of the distance function is dependent on the type of variables to be simulated; from our experience, the Manhattan distance is best suited for categorical variable while the weighted Euclidean distance works well for continuous variables, which is consistent with previous studies (Mariethoz et al 2010b); (3) in EnPAT, the pattern may include both categorical (i.e., rock facies) and continuous (i.e., piezometric head), in which case, the Manhattan distance should be used for the categorical variable and the weighted Euclidean distance for the continuous one, consequently, two distance tolerances should be defined, one for each type of data.

### 3 Synthetic Examples

#### 3.1 Reference Field and Cases

A single phase transient groundwater model is used to demonstrate the performance of the proposed method. The model is discretized into  $50 \times 50 \times 1$  grid cells with cell size  $1 \text{ m} \times 1 \text{ m} \times 1 \text{ m}$ . Two types of conductivity images will be used, one will be categorical just distinguishing between high and low conductivity facies, and the other one will be continuous, in which the categorical field generated before is filled with continuous conductivities drawn from a multiGaussian distribution. For the categorical variable simulation, the reference facies is generated by the SNESIM code using the training image from Strebelle (2002) (Fig. 4A), and constant conductivity values equal to 10 and  $10^{-4} \text{ m/d}$  are assigned to sand and shale, respectively (Fig. 4C). For the cases of continuous variable simulation, the reference logconductivity is obtained by populating the previous facies with the conductivities generated by sequential Gaussian simulation using the parameters listed in Table 1, which results in a bimodal conductivity distribution (Fig. 4D).

The reservoir is assumed to be confined and with no-flow boundary conditions on south and north, prescribed head values ( $h = 0 \text{ m}$ ) on the western and eastern boundaries (Fig. 4B). A pumping well is located at the center of the reservoir with constant flow rate  $25 \text{ m}^3/\text{d}$ . The initial head values are  $0 \text{ m}$  over the whole domain. Specific storage is assumed constant and equal to  $0.01 \text{ m}^{-1}$ . Head data are simulated for two time steps of 1.3 days and 3.8 days and collected at nine wells. The simulated head data collected at the nine wells as well as the conductivity data at the same nine wells will be used as the conditioning data (Fig. 4C and E for the categorical conductivities, and Fig. 4D and F for the continuous one). The code MODFLOW (Harbaugh et al 2000) is utilized to solve the transient groundwater flow problem.

Nine cases are considered, the characteristics of which are shown in Table 2. The cases are designed to investigate the impact of using different types of conditioning information. More specifically, the impact of using only conductivity conditioning data, only piezometric head conditioning data, or both conductivity and piezometric conditioning data are analyzed. As a reference, the case for which no conditioning data are used is also analyzed. Cases 1 through 4 correspond to the categorical conductivities, and cases 5 through 8 to continuous conductivities. Case 9 is set to demonstrate the proposed approach conditional on multiple time step piezometric head data (i.e., jointly conditional on  $K$ ,  $h_{t=1.3}$  and  $h_{t=3.8}$ ).

For all the cases, 2500 unconditional conductivity realizations are generated by SNESIM as the initial fields for the data assimilation. The ensemble size in this case is rather large in order to obtain the best results for validating the approach. Later on, in Sect.4 some approaches for reducing the size of the ensemble are presented and some of these have also been implemented in a paper by Li et al (2013). The maximum search radius to build the conditioning pattern is set at  $30 \text{ m}$  for both conductivity and head, and the maximum number of elements in the pattern is fixed at 10. The distance tolerance for categorical conductivity is 0 and for continuous conductivity is 0.01, and for the head it is specified to be 0.00005. The specific distance function used for each case and for each variable is given in Table 2. This set of parameters in EnPAT are obtained by sensitivity analysis.

### 3.2 Performance Evaluation

The performance of the proposed method in the synthetic examples is assessed using the following criteria:

1. From a qualitative point of view, the following items are evaluated:
  - the ensemble mean map: to check the reproduction of the main structures of the reference.
  - the ensemble variance maps: to visualize the spatial uncertainty of the estimation.
2. From a more quantitative point of view, two metrics are calculated:
  - Average absolute error (*AAE*). *AAE* measures the accuracy of the estimation, and is defined as

$$AAE(x) = \frac{1}{n} \sum_{i=1}^n (\bar{x}_i - x_{\text{ref},i}), \quad (5)$$

where  $x_i$  is either log-conductivity  $\ln K$  or hydraulic head  $h$  at location  $i$ ,  $\bar{x}_i$  represents its ensemble mean value at node  $i$ ,  $x_{\text{ref},i}$  is the reference value at location  $i$ ,  $n$  is the number of nodes.

- Average ensemble spread (*AES*). *AES* indicates the uncertainty of the estimation and can be written as

$$AES(x) = \frac{1}{n} \sum_{i=1}^N \sigma_{x_i}^2, \quad (6)$$

where  $\sigma_{x_i}^2$  is the ensemble variance at location  $i$ .

3. The main purpose of inverse conditioning is to make a better prediction of the future performance of the reservoir. Thus, the predictive capability of updated models will be assessed as discussed in Sect. 3.4.

### 3.3 Simulation Results

Ensembles of conductivity and piezometric head are generated using the EnPAT algorithm. Note that, for cases 1, 2, 5, and 6, piezometric head is obtained by running the flow simulation on the conductivity fields generated, and for the other cases, the head is directly simulated by EnPAT. Figure 5 shows the sand probability and variance for the cases 1, 2 and 4. Figure 6 displays the ensemble mean and variance of the log-conductivity for the cases 5, 6 and 8. Figure 7 and 8 show the ensemble mean and variance of simulated head. Figure 10 shows randomly selected individual realizations of conductivity and head for some of the cases. Figure 9 displays the sand probability and variance and the ensemble mean and variance of simulated head for the case 9. The calculated metrics *AAE* and *AES* are listed in Table 3. For the categorical variable simulation (i.e., cases 1 through 4),  $AAE(K)$  and  $AES(K)$  are calculated on the basis of the indicator transform of the conductivities (1 for the high values, 0 for the low ones).

### 3.3.1 Conditioning on Conductivity only

For cases 1 and 5, the unconditional facies fields are generated by the SNESIM code and further populated with conductivities for each facies. As expected, the sand probability and ensemble variance are almost constant values in the spatial distribution because the channels are randomly distributed across the realizations since there are no conditioning to any data (Fig. 5 and Fig. 6).

For cases 2 and 6, the conductivity fields are simulated using EnPAT conditioned to the measured  $K$  data at the nine conditioning wells. The realizations generated from case 1 and 5 are used as the ensemble training images. (Notice that for these two cases, the original facies image could have been used to generate each of the ensemble members, since the non-stationarity problem associated to conditioning to piezometric heads is not present here.) The results show that, as in the traditional MPS method, (1) measured data are exactly conditioned with zero variance at measurement locations; (2) the sand patterns disclosed by the probability maps (for the categorical conductivities) or by the mean maps (for the continuous conductivities) are close to the sand channels in the reference; (3) the individual realizations in Fig. 10 show channel features that are similar to the ones in the training images; (4) by analyzing the two metrics from Table 3, it is evident that both  $AAE$  and  $AES$  become smaller after conditioning to the measured conductivities than for the unconditional cases.

### 3.3.2 Conditioning on Piezometric Head only

Unlike the traditional MPS methods commonly focusing on the estimation of geologic parameters such as conductivity, in EnPAT the state (in our case piezometric head) is also simulated using the pattern search algorithm EnPAT. Because head patterns are affected by boundary conditions and the presence of source or sinks, the pattern search is implemented locally across the ensemble by analyzing only the node being simulated in all realizations. No conductivity fields are simulated in these cases, just the piezometric head fields, conditioned to the measured values at the 9 wells and using the piezometric head training images calculated from the conductivity realizations in cases 1 and 5.

For cases 3 and 7, the ensemble mean and variance of the simulated heads are displayed in Fig. 7 and 8. The results show that: (1) measured heads are honored exactly in both cases with zero variance at measurement locations; (2) the ensemble mean of piezometric head displays a pattern similar to the reference head in Fig. 4; (3) compared to the results from the cases conditioned only to conductivity, conditioning to pattern of piezometric head drastically reduces the uncertainty in the simulated head; (4) the simulated head fields reflect the patterns of variability exhibited by the corresponding conductivity fields; (5) Table 3 shows that  $AAE(h)$  and  $AES(h)$  both have a remarkable reduction compared to cases without any conditioning;

### 3.3.3 Conditioning to Both Conductivity and Piezometric Head Data

Cases 4, 8 and 9 are designed to demonstrate the capability of EnPAT to condition to conductivity and head data jointly, and also to jointly generate conductivity and head fields.

By analyzing the ensemble mean and variance maps in Fig. 5 to 9, the results show that: (1) both the measured conductivity and head data are exactly honored; (2) the pattern of ensemble mean maps is closer to the reference than for the cases with only conditioning to conductivity; (3) the ensemble variance of conductivity is reduced with respect to conditioning only to conductivity if the simulation is further conditioned to head data; (4) the simultaneously simulated conductivity and head fields exhibit patterns close to the patterns observed in the reference fields; (5) some noise is observed in the simulated results (Fig. 10) but is considered acceptable because a stronger constraint in the form of correlation between the conductivity and head is enforced. It can be further reduced by either increasing the number of ensemble members or reducing the distance tolerance values; (6) the two metrics  $AAE$  and  $AES$  have the lowest values, indicating that the highest accuracy and lowest uncertainty are achieved after conditioning to both measured conductivity and head values; (7) in Fig 9, the pattern exhibiting in the sand probability of case 9 is closer to the reference than the case with only conditioning one time step piezometric head; Also the simulated head is close to the head simulated from the reference displayed in Fig. 11; (8) It is noted in Table 3 that the value of  $AES$  for the case with conditioning to head data only is smaller than that where the conditioning is to both conductivity and head data. This is because when both head and conductivity are used as conditioning data, the pattern search is much more challenging and so the tolerances for the distances have to be relaxed. Despite this, the difference in  $AES$  between the two cases is quite small.

In conclusion, the EnPAT algorithm can be used to simultaneously generate conductivity and head realizations using a multiple-point statistics approach and therefore is able to capture features that cannot be captured by methods solely based on the characterization of the spatial variability by a two-point covariance.

### 3.4 Flow Prediction

Flow predictions are conducted for the first 8 cases for one time step. Using the conditional realizations of both conductivity and piezometric head for all 8 cases, piezometric head was predicted for  $t = 3.8$  days.

Figure 11 and 12 display the head fields corresponding to the reference data, and the ensemble mean and variance of head fields corresponding to cases 2, 4, 6, and 8. The unconditional cases 1 and 5 are not been displayed, and cases 3 and 7 have not been analyzed since for these two cases only the piezometric head realizations were generated (no update of the conductivity fields was performed). Table 3 shows the calculated metrics for the predicted head at the  $t = 3.8$  days. The results indicate that: (1) the accuracy of the predicted heads improves both for the categorical and continuous conductivity fields as conditioning data are taken into account; (2) similarly, the uncertainty diminishes as conditioning data are incorporated; (3) the improvement in prediction accuracy for the categorical conductivity cases is more significant than for the continuous cases when using the conditioning head data. This is because the conditioning head data from the categorical case reflect the channel structures better than in the continuous case (Fig 4E and F) because of the particular contrast between the channel and the mudstone assumed for these cases.

## 4 Discussion

A new inverse conditioning approach was proposed and demonstrated to simultaneously simulate parameter and state variables in a non-multiGaussian setting. The methodology extends the DS method for multiple-point pattern simulation and offers the capability to achieve incremental updating of reservoir models in real-time.

The inverse conditioning both for the measured conductivity and head data, in the example, is carried out as a one-step procedure. A drawback of this joint simulation, based on locally matching conditioning patterns through an ensemble of realizations is that a large ensemble size is required for the matches to be found. In practice, the computational cost could be reduced in the following aspects: (1) the measured conductivity data can be conditioned with traditional fast MPS method first followed by the conditioning to state data. This could potentially reduce the size of the ensemble required; (2) patterns here are only searched at the exact same location as the node simulated, which makes it more difficult to find the matched pattern. The location could be extended to a small neighborhood of simulated node as implemented in Zhou et al (2012) and Li et al (2013), which could reduce the number of training realizations required; (3) the pattern search scheme can be coupled with the pilot points concept as implemented in the sequential self-calibration (Gómez-Hernández et al 1997). Specifically, the pattern search is applied on the predefined pilot points with constraints of both  $K$  and  $h$ , and then a fast MPS method (i.e., only constraint of  $K$ ) is used to complete the simulation conditional on the pilot points. Li et al (2013) used a synthetic example to demonstrate that the computational cost is reduced and the corresponding updated parameter are improved with regard to the quality; (4) As done in DS by Rezaee et al (2013), EnPAT could also be extended to past a bunch of node at a time instead of only a single point when the matched pattern is found; (5) a parallel implementation of the EnPAT algorithm is straightforward, both for the flow simulations and for the generation of the conductivity and state variables (Tahmasebi et al 2012b).

EnPAT is initially inspired by the EnKF real-time data assimilation algorithm (Evensen 2003). As in EnKF, both the parameter and state variables are continuously updated, and thus it is well suitable for real-time updating using incremental data. In both methods, the correlation between parameter and state variables is explicitly estimated through the ensemble. A spurious correlation could be observed when the ensemble size is small and/or the transfer function is extremely non-linear such as multiple phase flow, which results in the inconsistency between the updated state and parameter (i.e., it does not honor the flow equation). Two approaches are commonly used to resolve this issue: (1) the most straightforward way is to increase the ensemble size, however the high computational cost of flow simulations could be a deterrent; A fast proxy of flow simulator coupling with the data assimilation algorithm could improve the results (He et al 2013); (2) only the geologic parameter is updated, and the corresponding state is obtained by running the forward simulator based on the estimated parameter so that the consistency is ensured; this approach has already been used in the EnKF-based method (Wen and Chen 2006; Jafarpour and Khodabakhshi 2011; Hu et al 2012) and in EnPAT (Li et al 2013). Here, in the synthetic example, as shown in Fig. 13, the simulated head field from EnPAT is consistent with the head simulated from MODFLOW

using the updated conductivity from the EnPAT for the two randomly selected individual realizations in the case of categorical conductivity. Recall that, here the pattern is matched through a large ensemble size of training images. An extensive comparison between the current implementation of EnPAT and the one by running forward simulation to derive the state variable is expected in the future work, including the impact of the ensemble size of training images and the evolution of simulated head when new data are conditioned.

Also, similar to EnKF, the ensemble updated parameter and state can be used for the prediction, and the uncertainty can be assessed by running the flow simulation on the corresponding ensemble. Whether the uncertainty after the data conditioning could represent the posterior probability remains further investigation. For example, the uncertainty could be evaluated by comparing the EnPAT with rejection sampling.

A key criticism against EnPAT could be that the method needs the flow response corresponding to an ensemble of models in order to search for the pattern of state variables such as head and this could be CPU-time intensive. However, this is not the case, as the flow simulation has only to be performed for one time-step using the updated state variables from the previous time step as the starting state. This is exactly the same as implemented in EnKF.

## 5 Conclusions

With the increasing use of MPS methods to characterize complex geological formations, developing an innovative inverse method by integrating dynamic data within the MPS framework is important. In this paper, a novel inverse method (termed EnPAT) was proposed to handle data assimilation within the MPS framework. Parameters and state variables are simultaneously estimated using a pattern search approach. The method offers the additional advantage of incremental data assimilation as in EnKF. The important characteristics of the pattern simulation algorithm (EnPAT) are: (1) parameter and state variables do not have to be model by multiGaussian distributions; (2) the relationship between parameters and state variables do not have to be linear; (3) the algorithm can use real-time data from on-line sensors in order to perform incremental updating of reservoir models.

The EnPAT algorithm is demonstrated for both categorical and continuous conductivity fields by conditioning to the observed conductivity and head data. The results show that the measured conductivity and head data can be honored in the updated model and flow predictions are more accurate, and more importantly, curvilinear geologic structures are preserved after data integration.

**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 Science and Innovation through project CGL2011-23295. The authors also wish to thank the guest editors Philippe Renard and Grégoire Mariethoz as well as three anonymous reviewers for their comments, which substantially helped improving the final version of the manuscript.

## References

- Aanonsen S, Nævdal G, Oliver D, Reynolds A, Valles B (2009) The ensemble kalman filter in reservoir engineering—a review. *SPE J* 14(3):393–412
- Abdollahifard MJ, Faez K (2013) Stochastic simulation of patterns using bayesian pattern modeling. *Computat Geosci* 17(1):99–116
- Alcolea A, Renard P (2010) Blocking Moving Window algorithm: Conditioning multiple-point simulations to hydrogeological data. *Water Resour Res* 46:W08511
- Arpat GB (2005) Sequential simulation with patterns. PhD thesis, Stanford University
- Caers J (2002) Geostatistical history matching under training-image based geological model constraints. *SPE Annual Technical Conference and Exhibition, SPE* 77429
- Caers J (2003) Efficient gradual deformation using a streamline-based proxy method. *J Pet Sci Eng* 39(1-2):57–83
- Chen Y, Zhang D (2006) Data assimilation for transient flow in geologic formations via ensemble Kalman filter. *Adv Water Resour* 29(8):1107–1122
- Evensen G (2003) The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynam* 53(4):343–367
- Fu J, Gómez-Hernández JJ (2009) A Blocking Markov Chain Monte Carlo Method for Inverse Stochastic Hydrogeological Modeling. *Math Geosci* 41(2):105–128
- Gómez-Hernández JJ, Sahuquillo A, Capilla JE (1997) Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data, 1, Theory. *J Hydrol* 203(1-4):162–174
- Gu Y, Oliver D (2006) The ensemble Kalman filter for continuous updating of reservoir simulation models. *J Energy Resour Technol* 128(1):79–87
- Guardiano F, Srivastava R (1993) Multivariate geostatistics: beyond bivariate moments. In: Soares A (ed) *Geostatistics-Troia*, Kluwer Academic Publ, Dordrecht, pp 133–144
- Hamill T, Whitaker J, Snyder C (2001) Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon Weather Rev* 129:2776–2790
- Harbaugh AW, Banta ER, Hill MC, McDonald MG (2000) MODFLOW-2000, the U.S. Geological Survey modular ground-water model. U.S. Geological Survey, Branch of Information Services, Reston, VA, Denver, CO
- He J, Sarma P, Durlofsky LJ (2013) Reduced-order flow modeling and geological parameterization for ensemble-based data assimilation. *Comput Geosci* 55:54–69
- 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 Resour Res* 44(9):W09408
- Hoffman BT, Caers J (2005) Regional probability perturbations for history matching. *J Petrol Sci Eng* 46(1-2):53–71
- Hu LY, Chuginova T (2008) Multiple-point geostatistics for modeling subsurface heterogeneity: A comprehensive review. *Water Resour Res* 44(11):W11413
- Hu LY, Zhao Y, Liu Y, Scheepens C, Bouchard A (2012) Updating multipoint simulations using the ensemble kalman filter. *Comput Geosci* 51:7–15
- Huang Y, Srinivasan S (2012) Efficient conditional simulation of spatial patterns using a pattern-growth algorithm. *Geostatistics Oslo 2012* pp 209–220
- Jafarpour B, Khodabakhshi M (2011) A probability conditioning method (PCM) for nonlinear flow data integration into multipoint statistical facies simulation. *Math Geosci* 43(2):133–164
- Kashib T, Srinivasan S (2006) A probabilistic approach to integrating dynamic data in reservoir models. *J Pet Sci Eng* 50(3):241–257
- Li L, Zhou H, Hendricks Franssen HJ, Gómez-Hernández JJ (2012) Modeling transient groundwater flow by coupling ensemble kalman filtering and upscaling. *Water Resour Res* 48(1):W01537
- Li L, Srinivadan S, Zhou H, Gómez-Hernández J (2013) A pilot point guided pattern matching approach to integrate dynamic data into geological model. *Adv Water Resour*, doi:10.1016/j.advwatres.2013.10.008
- Mariethoz G, Renard P, Caers J (2010a) Bayesian inverse problem and optimization with iterative spatial resampling. *Water Resour Res* 46(11):W11530
- Mariethoz G, Renard P, Straubhaar J (2010b) The direct sampling method to perform multiple-point geostatistical simulations. *Water Resour Res* 46:W11536



- Meerschman E, Pirot G, Mariethoz G, Straubhaar J, Meirvenne M, Renard P (2013) A practical guide to performing multiple-point statistical simulations with the direct sampling algorithm. *Comput Geosci* 52:307–324
- Ramarao BS, LaVenue AM, de Marsily G, Marietta MG (1995) Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields, 1, Theory and computational experiments. *Water Resour Res* 31(3):475–493
- Rezaee H, Mariethoz G, Koneshloo M, Asghari O (2013) Multiple-point geostatistical simulation using the bunch-pasting direct sampling method. *Comput Geosci* 54:293–308
- Straubhaar J, Renard P, Mariethoz G, Froidevaux R, Besson O (2011) An improved parallel multiple-point algorithm using a list approach. *Math Geosci* 43(3):305–328
- Strebelle S (2002) Conditional simulation of complex geological structures using multiple-point statistics. *Math Geol* 34(1):1–21
- Sun AY, Morris AP, Mohanty S (2009) Sequential updating of multimodal hydrogeologic parameter fields using localization and clustering techniques. *Water Resour Res* 45(7):W07424
- Tahmasebi P, Hezarkhani A, Sahimi M (2012a) Multiple-point geostatistical modeling based on the cross-correlation functions. *Comput Geosci* 16(3):779–797
- Tahmasebi P, Sahimi M, Mariethoz G, Hezarkhani A (2012b) Accelerating geostatistical simulations using graphics processing units (gpu). *Comput Geosci* 46:51–59
- Wen X, Chen W (2006) Real-time reservoir model updating using ensemble kalman filter with confirming option. *SPE J* 11(4):431–442
- Zhang T, Switzer P, Journel A (2006) Filter-based classification of training image patterns for spatial simulation. *Math Geol* 38(1):63–80
- 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. *Adv Water Resour* 34(7):844–864
- Zhou H, Gómez-Hernández J, Li L (2012) A pattern-search-based inverse method. *Water Resour Res* 48(3):W03505

**Table 1** Parameters of the random functions describing the spatial continuity of the sand and shale log-conductivities.

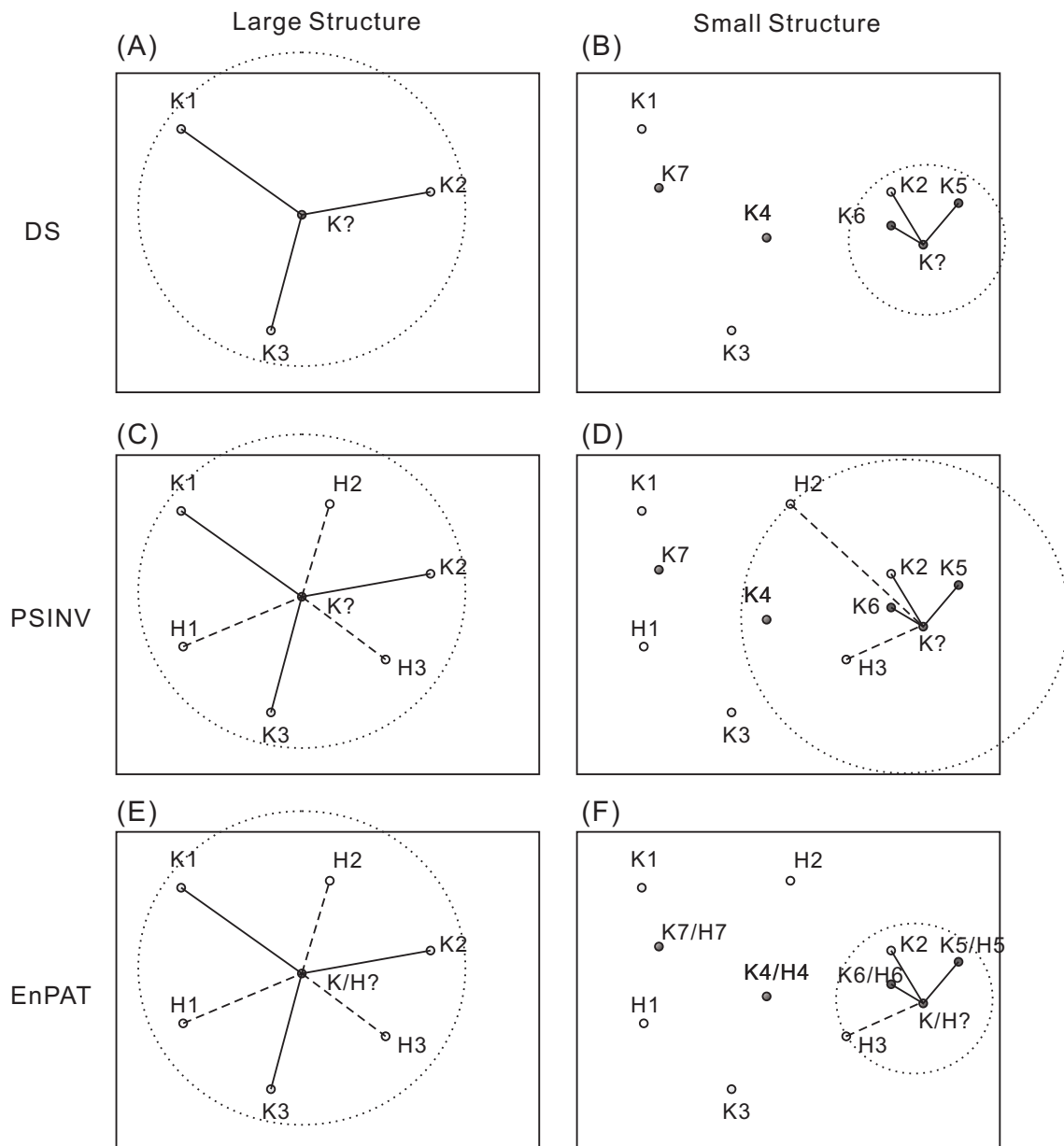
	Mean [ln m d <sup>-1</sup> ]	Standard deviation [ln [m d <sup>-1</sup> ]]	Variogram type	$\lambda_x$ [m]	$\lambda_y$ [m]	sill
Sand	2	1	exponential	60	30	1
Shale	-3	1	exponential	30	30	0.7

**Table 2** Definition of cases based on the different sets of conditioning data.

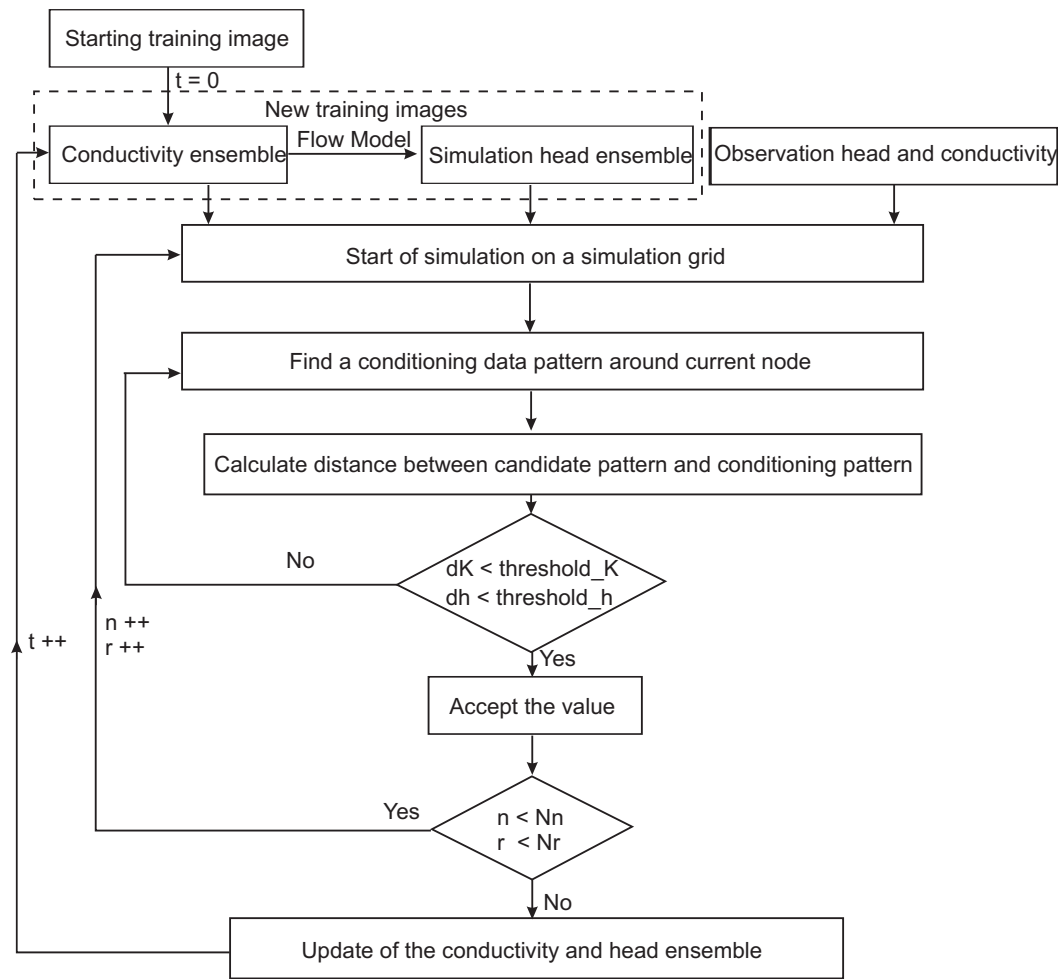
Case	Cate.( $K$ )	Cont.( $K$ )	Cond. $K$	Cond. $h_{t=1.3}$	Cond. $h_{t=3.8}$	Dist. func. ( $K$ )	Dist. func. ( $h$ )
1	×						
2	×		×			Manhattan	
3	×			×			Euclidean
4	×		×	×		Manhattan	Euclidean
5		×					
6		×	×			Euclidean	
7		×		×			Euclidean
8		×	×	×		Euclidean	Euclidean
9	×		×	×	×	Manhattan	Euclidean

**Table 3** AAE and AES of conductivity and head for the different cases. Notice that the values for  $t = 1.3$  are after updating the realizations using EnPAT, and the values for  $t = 3.8$  are for the predicted heads, prior to sampling new data and updating again the conductivity and head realizations.

Case	1	2	3	4	5	6	7	8
$AAE(K)$	0.422	0.269	-	0.239	2.184	1.513	-	1.471
$AES(K)$	0.210	0.155	-	0.123	6.234	4.363	-	3.789
$AAE(h_{t=1.3})$	1.088	0.294	0.233	0.225	0.422	0.143	0.169	0.107
$AES(h_{t=1.3})$	357.091	13.234	0.105	0.145	7.518	0.215	0.032	0.051
$AAE(h_{t=3.8})$	2.052	0.401	-	0.302	0.772	0.207	-	0.205
$AES(h_{t=3.8})$	1041.721	35.917	-	0.294	17.859	0.496	-	0.366



**Fig. 1** Spatial evolution of the pattern search for three MPS algorithms. The open circles denote conditioning data and solid circles simulated values which are incorporated to the conditioning data set after having being simulated. PSINV is the acronym for the pattern search inverse method developed by Zhou et al (2012).



**Fig. 2** Flowchart of the algorithm.  $N_n$  denotes the number of nodes in the domain,  $N_r$  corresponds to the number of realizations simulated.

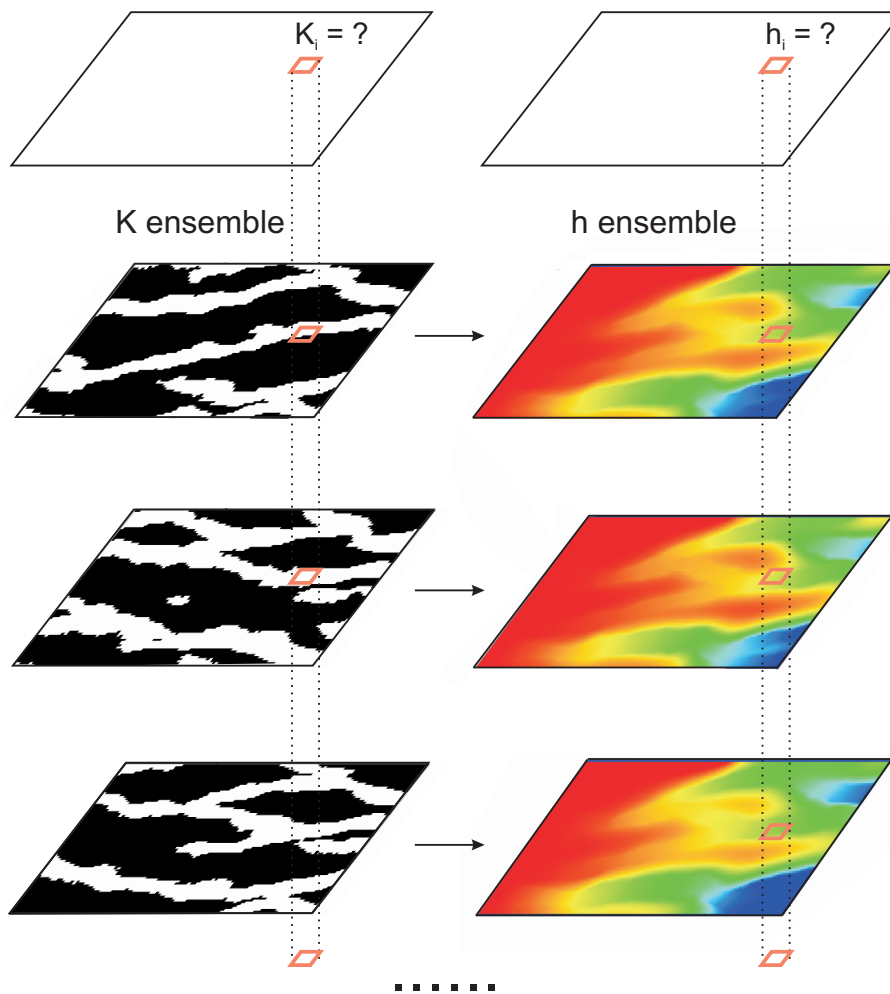
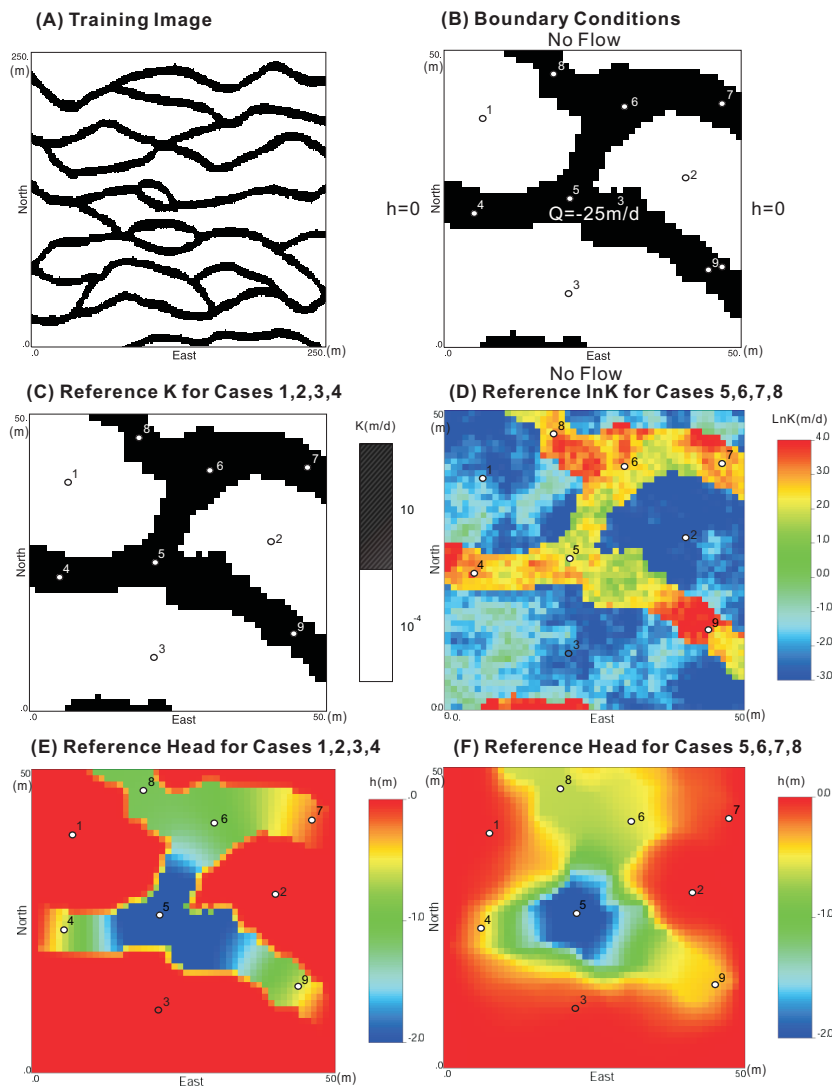
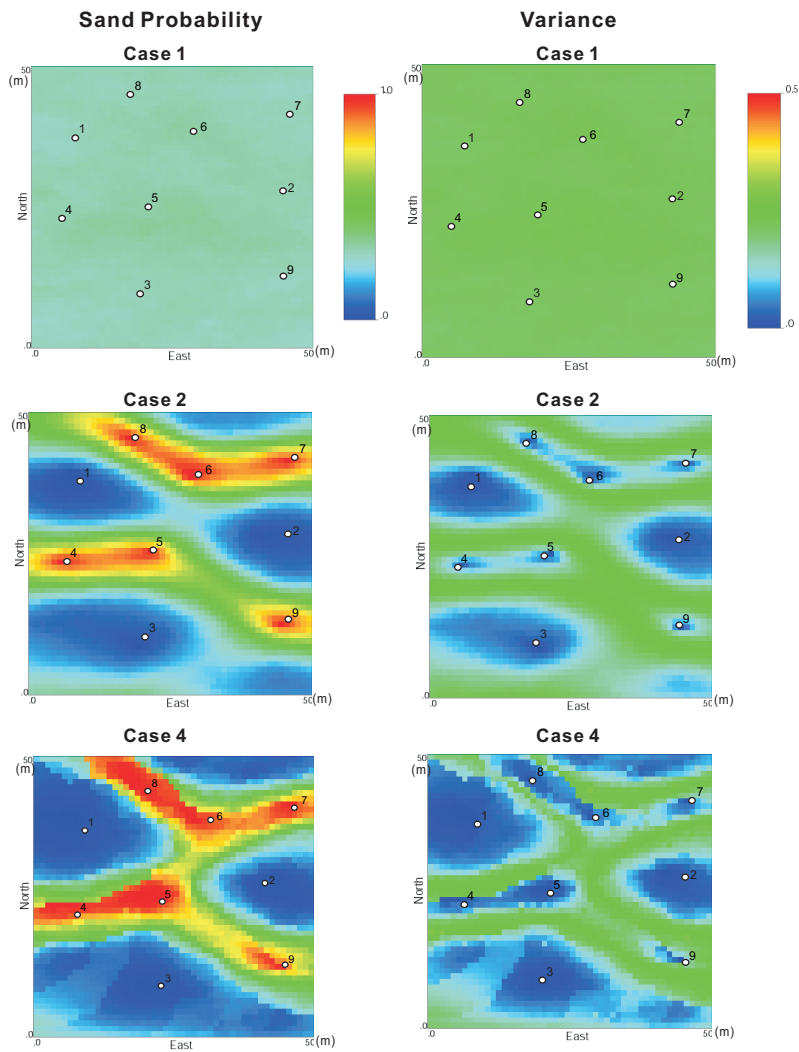


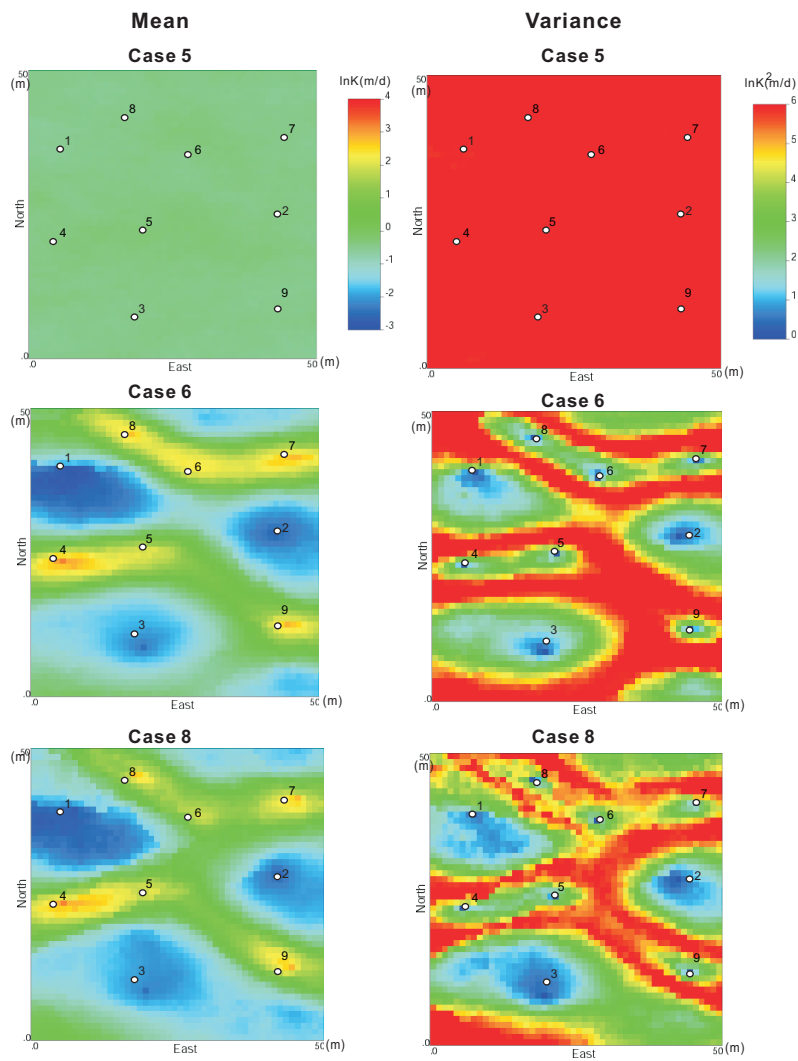
Fig. 3 Pattern search with ensemble training images.



**Fig. 4** (A) Training image. (B) Boundary condition of flow model and observation wells. (C) Reference  $K$  field for categorical simulations. (D) Reference  $\ln K$  field for continuous simulations. (E) Reference piezometric head field ( $h_{t=1.3}$ ) for cases 1, 2, 3 and 4. (F) Reference piezometric head field ( $h_{t=1.3}$ ) for cases 5,6,7 and 8

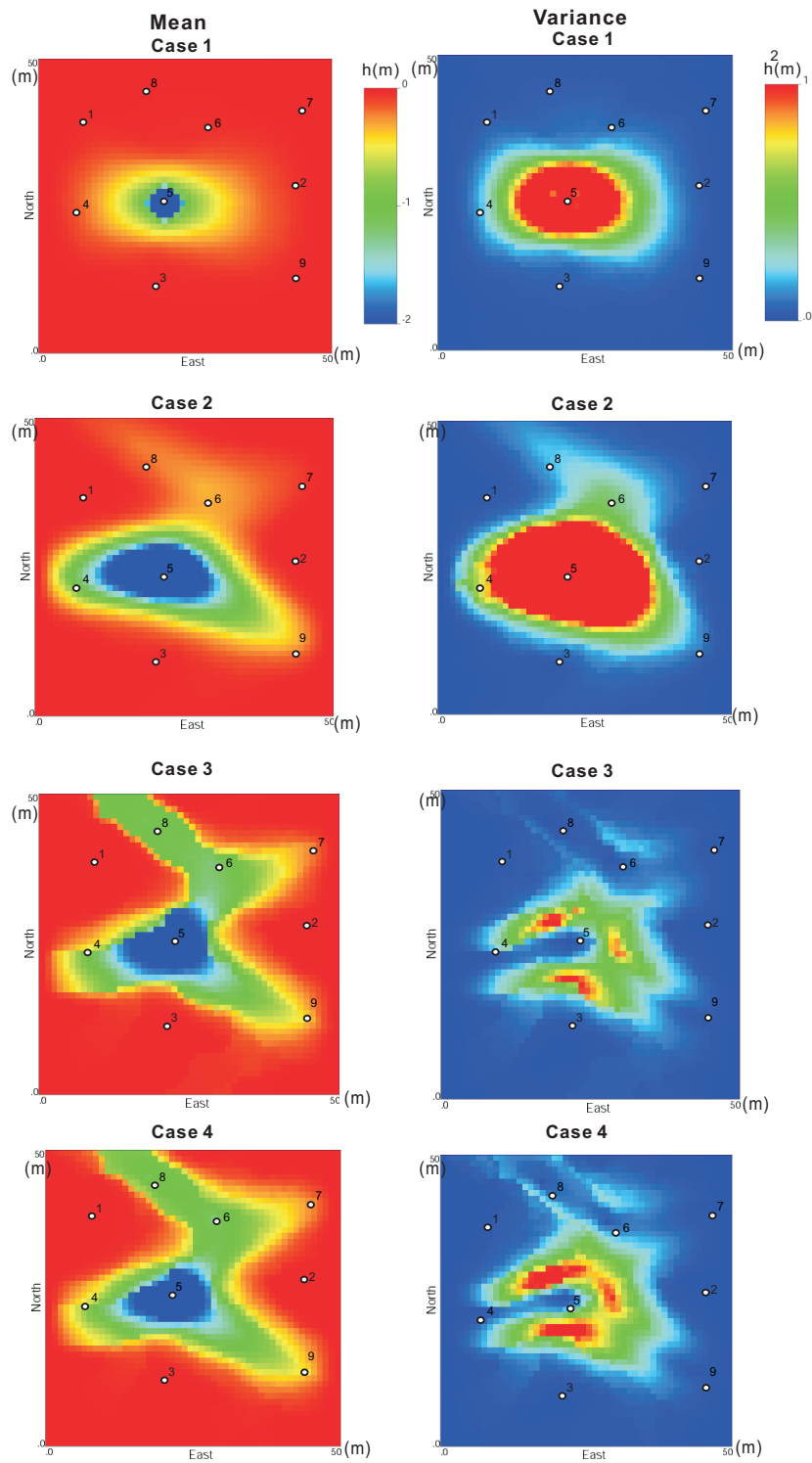


**Fig. 5** Sand probability and variance of simulated categorical conductivity calculated from 2500 ensemble realizations. For case 1 (without any conditioning), the conductivity is obtained through the MPS method while for cases 2 (conditioning on  $K$ ) and 4 (jointly conditioning on  $K$  and  $h_{t=1.3}$ ), the conductivity is simulated by running EnPAT.

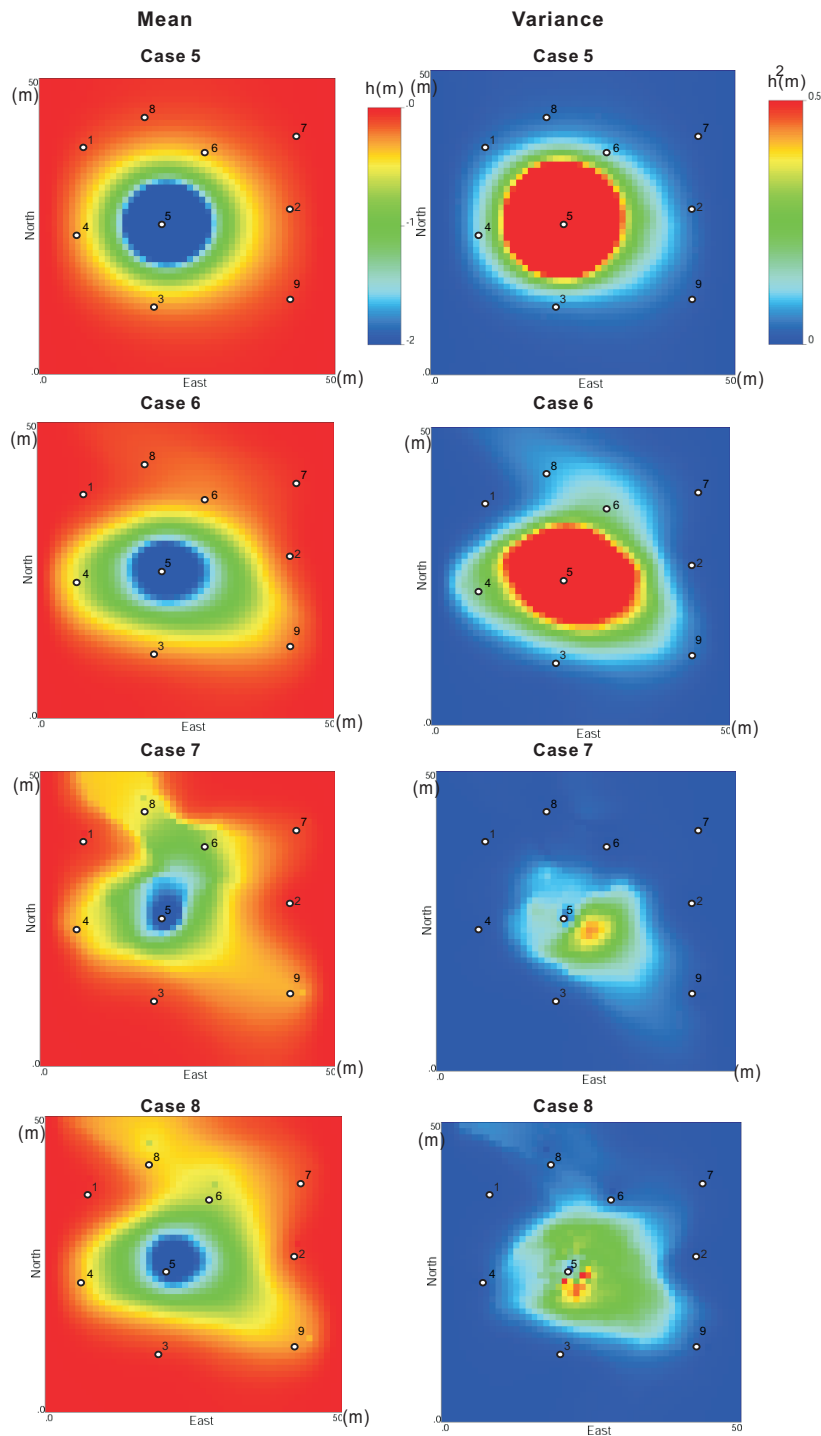


**Fig. 6** Mean and variance of simulated continuous conductivity. For case 5 (without any conditioning), the conductivity is obtained through the MPS method while for cases 6 (conditioning on  $K$ ) and 8 (jointly conditioning on  $K$  and  $h_{t=1,3}$ ), the conductivity is simulated by running EnPAT.

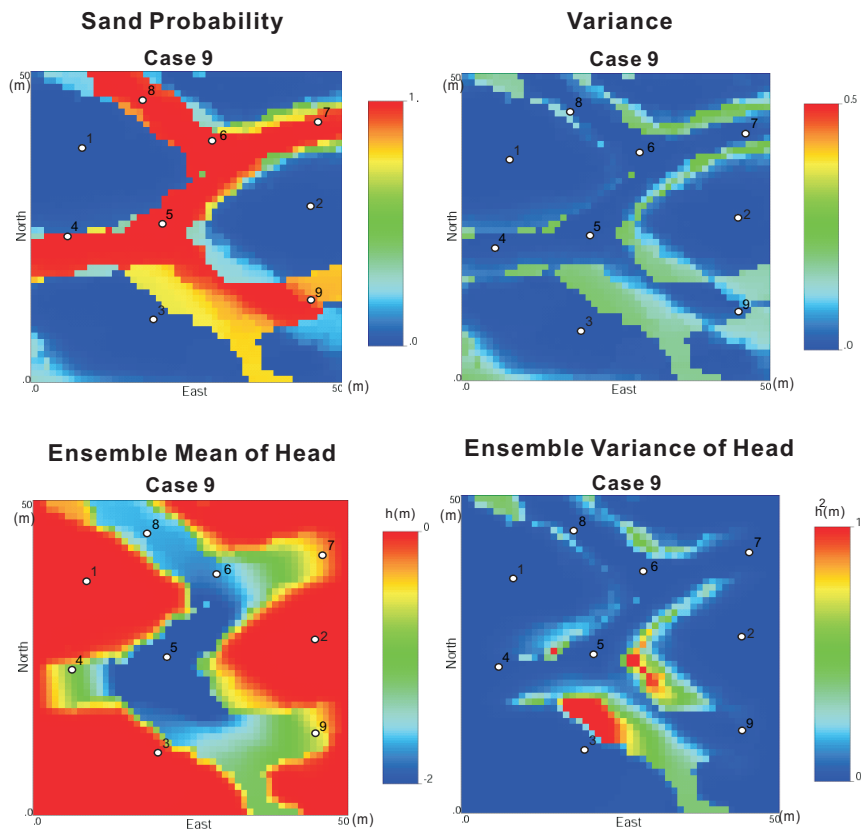




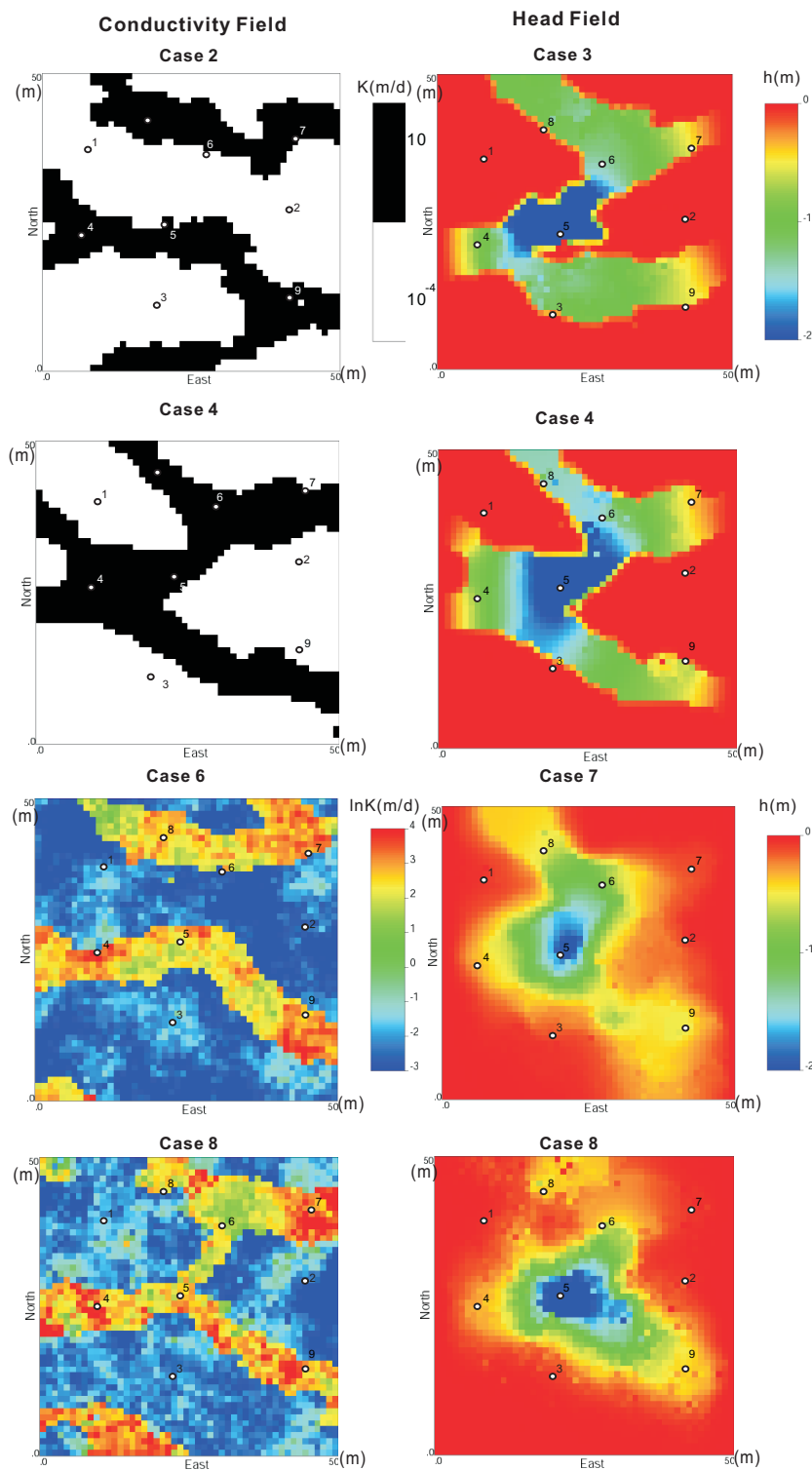
**Fig. 7** Mean and variance of simulated head for the categorical conductivity cases. For cases 1 (without any conditioning) and 2 (conditioning on  $K$ ), the head is obtained from the flow simulation while for cases 3 (only conditioning on  $h_{t=1.3}$ ) and 4 (jointly conditioned on  $K$  and  $h_{t=1.3}$ ), the head is simulated by running EnPAT.



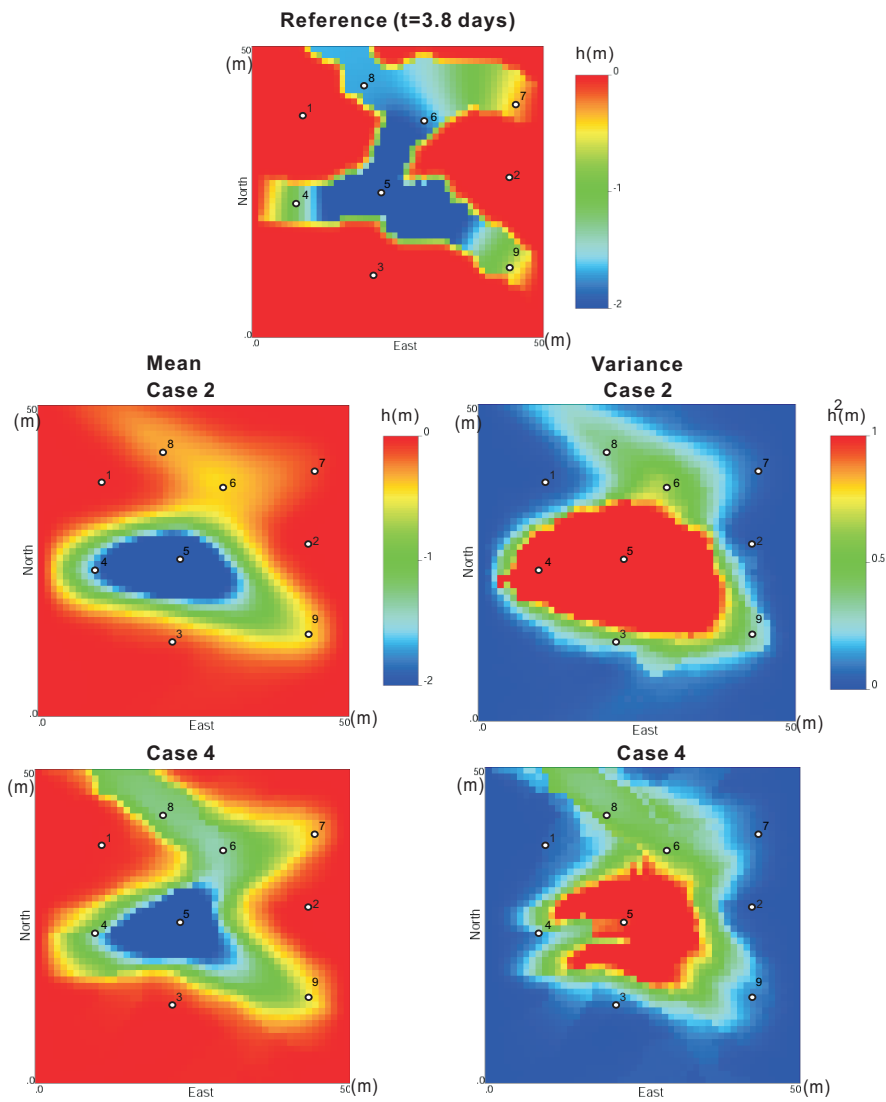
**Fig. 8** Mean and variance of simulated head for the continuous conductivity cases. For cases 5 (without any conditioning) and 6 (only conditioning on  $K$ ), the head is obtained from the flow simulation while for cases 7 (conditioning on  $h_{t=1.3}$ ) and 8 (jointly conditioning on  $K$  and  $h_{t=1.3}$ ), the head is simulated by running EnPAT.



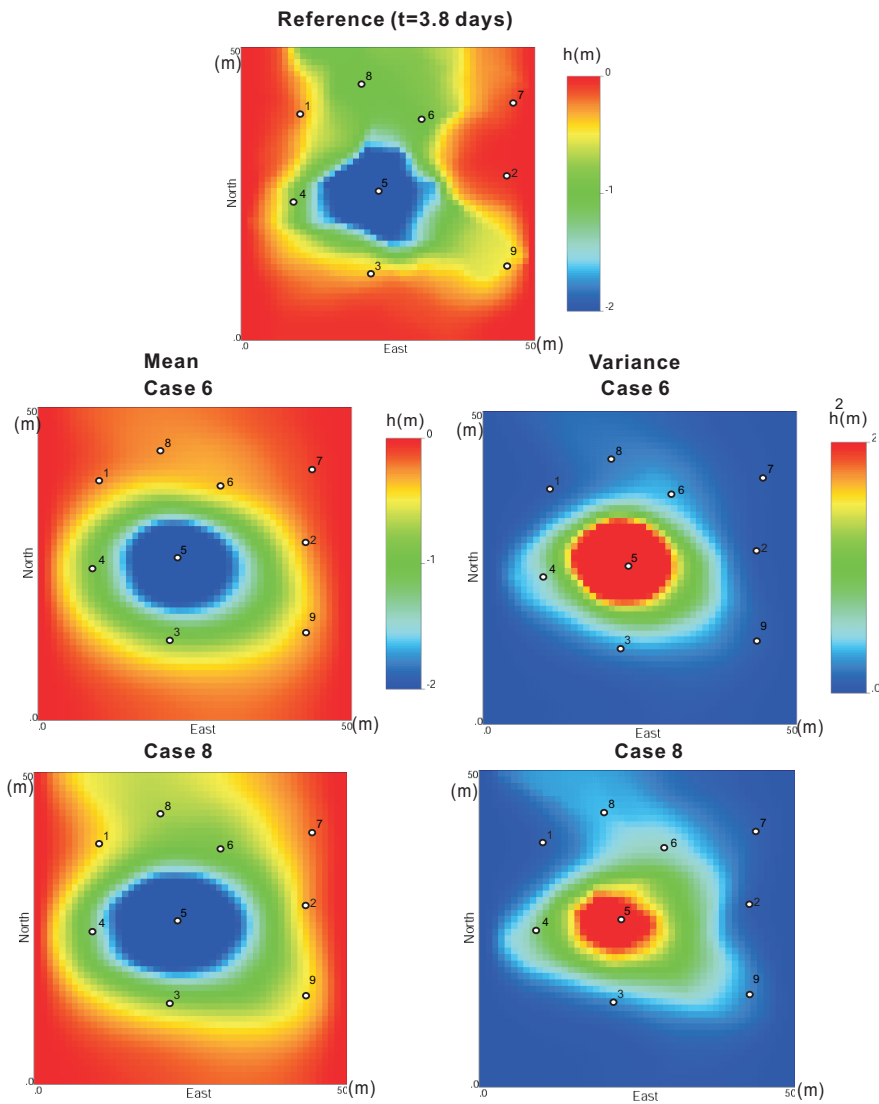
**Fig. 9** Sand probability and variance of simulated categorical conductivity and corresponding ensemble mean and variance of simulated piezometric head for case 9 (jointly conditioning on  $K$ ,  $h_{t=1.3}$  and  $h_{t=3.8}$ ).



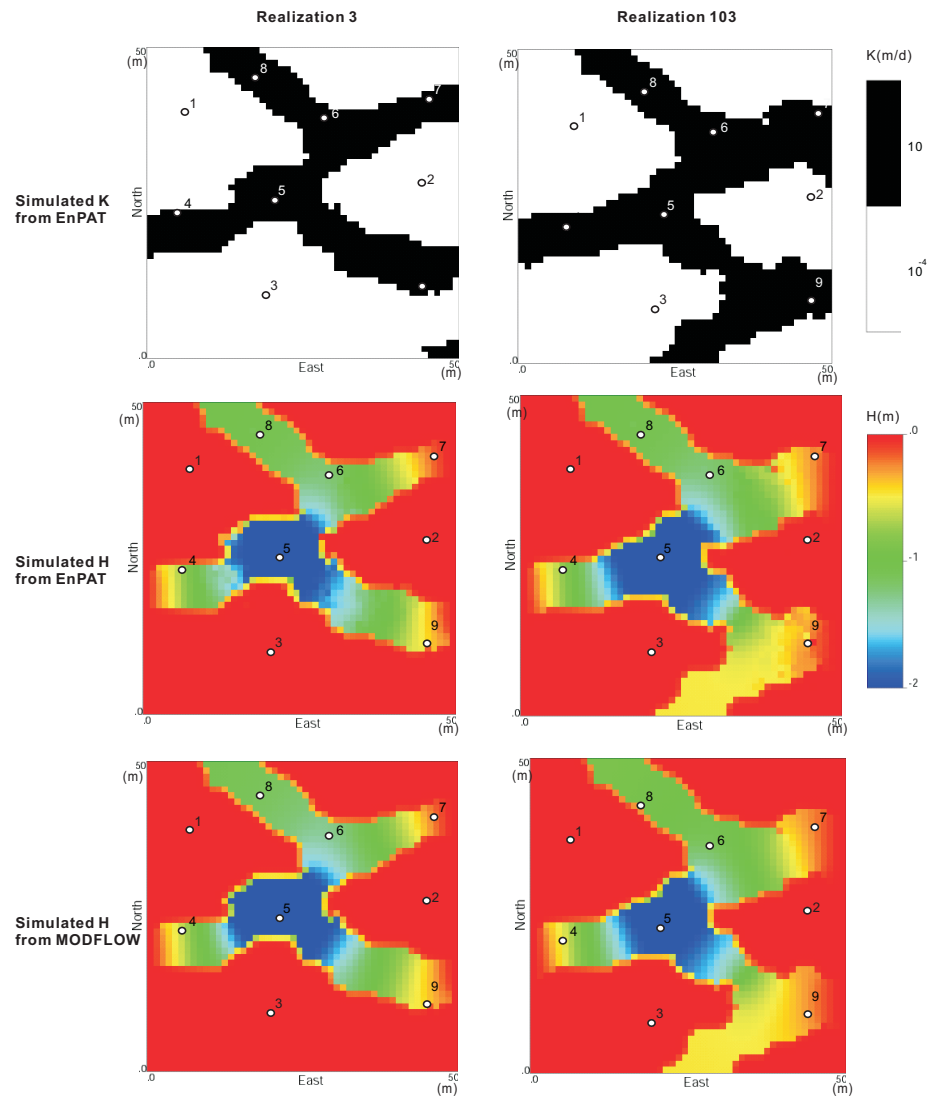
**Fig. 10** Randomly selected individual realizations simulated with EnPAT for cases 2 and 6 (conditioning on  $K$ ) and cases 4 and 8 (jointly conditioning on  $K$  and  $h_{t=1.3}$ )



**Fig. 11** Flow predictions for the categorical conductivity case 2 (conditioning on  $K$ ) and 4 (jointly conditioning on  $K$  and  $h_{t=1.3}$ ).



**Fig. 12** Flow predictions for the continuous conductivity case 6 (conditioning on  $K$ ) and 8 (jointly conditioning on  $K$  and  $h_{t=1.3}$ ).



**Fig. 13** Comparison of simulated head ( $h_{t=1.3}$ ) between EnPAT and MODFLOW for the categorical conductivity case