

Document downloaded from:

<http://hdl.handle.net/10251/39704>

This paper must be cited as:

Li, L.; Srinivasan, S.; Zhou, H.; Gómez-Hernández, JJ. (2013). A Pilot Point Guided Pattern Matching Approach to Integrate Dynamic Data into Geological Modeling. *Advances in Water Resources*. 62(Part A):125-138. doi:10.1016/j.advwatres.2013.10.008.



The final publication is available at

<http://dx.doi.org/10.1016/j.advwatres.2013.10.008>

Copyright Elsevier

# A Pilot Point Guided Pattern Matching Approach to Integrate Dynamic Data into Geological Modeling

Liangping Li<sup>a,\*</sup>, Sanjay Srinivasan<sup>a</sup>, Haiyan Zhou<sup>a</sup>, J. Jaime Gómez-Hernández<sup>b</sup>

<sup>a</sup>Center for Petroleum and Geosystems Engineering Research, University of Texas at Austin, 78712, Austin, USA

<sup>b</sup>Research Institute of Water and Environmental Engineering, Universitat Politècnica de València, 46022, Valencia, Spain

---

## Abstract

Methods based on multiple-point statistics (MPS) have been routinely used to characterize complex geological formations in the last decade. These methods use the available static data (for example, measured conductivities) for conditioning. Integrating dynamic data (for example, measured transient piezometric head data) into the same framework is challenging because of the complex non-linear relationship between the dynamic response and geology. The Ensemble PATtern (EnPAT) search method was recently developed as a promising technique to handle this problem. In this approach, a pattern is postulated to be composed of both parameter and state variables, and then, parameter values are sequentially (point-wise) simulated by directly sampling the matched pattern from an ensemble of training images of both geologic parameters and state variables. As a consequence, the updated ensemble of realizations of the geological parameters preserve curvilinear structures (i.e., non-multiGaussianity) as well as the complex relationship between static and dynamic data. Moreover, the uncertainty of flow and transport predictions can be assessed using the updated ensemble of geological models. In this work, we further modify the EnPAT method by introducing the pilot-point concept into the algorithm. More specifically, the parameter values at a set of randomly selected pilot point locations are simulated by the pattern searching procedure, and then a faster MPS method is used to complete the simulation by conditioning to the previously simulated pilot point values. This pilot point guided MPS implementation results in lower computational cost and more accurate inference of the parameter field. In addition, in some situations where there is sparsity of measured geologic static data, the EnPAT algorithm is extended to work only with the dynamic data. We employed a synthetic example to demonstrate the effectiveness of pilot points in the implementation of EnPAT, and also the capability of dynamic data to identify complex geologic structures when measured static data are not available.

---

\*Corresponding author

Email addresses: liangpingli@utexas.edu (Liangping Li), sanjay.srinivasan@engr.utexas.edu (Sanjay Srinivasan), haiyanzhou@utexas.edu (Haiyan Zhou), jaime@dihma.upv.es (J. Jaime Gómez-Hernández)

*Keywords:* multiple-point geostatistics, conditional simulation, inverse modeling, ensemble-based methods, history matching

---

## 1. Introduction

Mathematical modeling of subsurface flow and transport is essential for managing energy production and contaminant remediation. Aquifer parameters such as hydraulic conductivity or permeability, exhibit large spatial variation commonly over several orders of magnitude. Due to the scarcity of measurements, a geostatistical approach [e.g., 1, 2, 3, 4, 5, 6] is usually employed to represent the spatial heterogeneity of aquifer attributes. These geostatistical models are conditioned to the measured static data and yield multiple equiprobable realizations of the attributes. The uncertainty of model response is assessed subsequently by running a forward model on these multiple parameter realizations [7, 8, 9]. Besides static data (i.e., the hard data), dynamic data such as transient piezometric head and concentration can also be used to condition the models. The procedure of constructing aquifer models conditioned to dynamic data is termed inverse modeling where the objective is to identify the parameter values at unsampled locations by integrating those dynamic data into the model, and thus to improve the predictions of flow and transport in the future [e.g., 10, 11, 12, 13, 14, 15].

Inverse methods have been developed and used extensively to generate permeability or hydraulic conductivity models conditioned to dynamic data [e.g., 16, 17, 18, 19, 20, 21]. De Marsily et al. [16] developed the pilot-point method, in which the a conductivity map is determined by calibrating a few pilot-point locations followed by a kriging interpolation. Gómez-Hernández et al. [19] proposed the self-calibration stochastic inverse method, which is an extension of the pilot point method, aimed at the generation of multiple conductivity realizations, all matching the observed state data. The performance of the self-calibration method has been demonstrated for synthetic and real case studies [e.g., 22, 23, 24]. One key concern of this approach is how to determine the number of pilot points and their locations. Gómez-Hernández et al. [19] recommended two or three pilot points per correlation length. LaVenue and Pickens [25] placed the pilot points in the highest sensitivity regions. Wen et al. [26] proposed to randomly locate the pilot points such that the spacing between the pilot points is one correlation length. Wen et al. [27] coupled self-calibration with genetic algorithms to determine the optimal locations of pilot points. A code implementing the self-calibration model is available to the public [28].

Beside the pilot point-based inverse methods, Hu [20] proposed the gradual deformation method, in which a single deformation parameter controls the generation of conductivity fields such that the simulated state values match the observation data. Evensen [21] proposed the ensemble Kalman filter, a further extension of the extended Kalman filter, in which the covariance between the aquifer attribute at a location and the corresponding well response is calculated from an ensemble of realizations and is used subsequently to

32 update the ensemble so as to reflect the measured well response. Heidari et al. [29] proposed to update the  
33 conductivity fields at the pilot point locations using the EnKF, and then extrapolate the updated values to  
34 all locations in the aquifer by kriging.

35 All the inverse methods mentioned above are optimal for multi-Gaussian geologic media. In other words,  
36 they perform well for conductivity fields following a multiGaussian random function such as those generated  
37 by two-point variogram-based geostatistical methods such as sequential Gaussian simulation [3]. However,  
38 traditional two-point covariance methods can not be used to describe fluvial depositions, which commonly  
39 display features of curvilinear geometry. The significance of a curvilinear feature on the flow and transport  
40 predictions has been discussed in the literature [30, 31, 32, 33, 34]. An alternative to two-point covariance  
41 methods is to use recently developed methods based on multiple-point statistics (MPS) to address this  
42 problem [4]. In this approach, instead of using the traditional variogram model, a training image that  
43 conceptually describes the salient geological features is used. A spatial template (i.e., a multiple-point  
44 configuration) is used to infer the experimental local conditional distributions [35]. A complete review of the  
45 training image based MPS method for aquifer modeling is presented in Hu and Chugunova [36]. Alternative  
46 approaches are available to generate non-multiGaussian field such as transition probabilities and copula  
47 methods [e.g., 37].

48 Inverse methods developed to work in conjunction with multiple-point-based simulation methods are  
49 relatively new in the literature [e.g, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48]. Caers [38] applied a probability  
50 perturbation method on the permeability fields generated from MPS. Alcolea and Renard [39] developed a  
51 block moving-window algorithm, an extension of the block Markov Chain Monte Carlo method by Fu and  
52 Gómez-Hernández [49] to condition MPS simulations to piezometric head data as well as connectivity data.  
53 Mariethoz et al. [40] proposed an iterative spatial resampling method working on the MPS simulation in the  
54 Bayesian framework. More recently, the ensemble Kalman filter (EnKF) is gaining popularity in petroleum  
55 engineering and hydrogeology because of its computational efficiency and real-time data assimilation features  
56 [e.g., 50, 51, 52, 53, 54, 55]. However, the classical EnKF, when implemented on models constrained to  
57 MPS, does not preserve the statistics or complexity of the models, because of the covariance-based updating  
58 used in the EnKF. The complex relationship between the spatial pattern of state variables and the flow  
59 response is approximated by a covariance, and the repeated updating results in a final ensemble that does  
60 not correctly exhibit the complex spatial connectivity of features such as channels, fractures, etc. Some  
61 variations of the EnKF have been proposed to overcome the limitation posed by the two-point covariance-  
62 based updating so that the connectivity of the permeability field is properly preserved. Sun et al. [41]

63 proposed to couple the EnKF and a mixture of Gaussian models as well as localization techniques in order  
64 to improve the performance for fluvial models. Sarma and Chen [42] developed a kernel EnKF to preserve the  
65 connectivity of the MPS-based permeability realizations. Zhou et al. [43, 56] proposed to first transform the  
66 parameter and state variables to marginal Gaussian distributions through the normal-score transformation  
67 before implementing the updating step in EnKF. Jafarpour and Khodabakhshi [44] suggested updating the  
68 ensemble mean of several MPS-based permeability realizations using EnKF and subsequently use the updated  
69 ensemble-mean values as soft data to regenerate updated models using the MPS approach. Hu et al. [45]  
70 proposed to update the uniform-score random numbers used to draw outcomes from the MPS conditional  
71 distributions, using the EnKF.

72 All the above-mentioned EnKF-based updating schemes may yield suboptimal representation of perme-  
73 ability variations in the aquifer because of the linearization of the transfer function model implied by the  
74 representation of the complex relationship between state and dynamic response variables using lower order  
75 moments (e.g., mean and covariance). Zhou et al. [47] proposed a pattern-search-based inverse method where  
76 the relationship between the conductivity field and the dynamic responses, in the form of corresponding pat-  
77 terns, is inferred from training images and used to simulate MPS-based conductivity fields. This process  
78 was subsequently extended in Li et al. [48] where an Ensemble PATten (EnPAT) algorithm was presented  
79 to integrate dynamic data within an ensemble-based multiple-point statistic framework. In this approach,  
80 model parameter and state values are simultaneously and sequentially estimated, which not only improves  
81 the characterization of the parameter field, but also makes it feasible to assimilate dynamic data in real-time,  
82 similar to other ensemble-based filtering approaches.

83 In this work, we further improve the performance of the EnPAT algorithm [48] by implementing the pilot-  
84 point concept as done in the sequential self-calibration method [16, 19]. More specifically, the conductivities  
85 at pilot point locations are generated through the EnPAT scheme, and then a fast MPS method is used to  
86 generate updates of the initial ensemble conditioned on the pilot-point parameter values. We demonstrate  
87 this algorithm on a synthetic data set. Moreover, in some cases, hard data (i.e., conductivity values used to  
88 condition the conductivity realizations) may be unavailable and only well responses may exist to generate  
89 the ensemble of aquifer models. We extend the EnPAT algorithm to condition only on the dynamic data  
90 in order to recognize curvilinear geologic structures. Lastly, we demonstrate the algorithm for conditioning  
91 to fully transient flow response information. In these demonstrations, we track the evolution of models as  
92 dynamic data is integrated sequentially in time.

93 The rest of the paper is organized as follows. Section 2 outlines the improved EnPAT methodology

94 and details the significance of the use of pilot points in this new algorithm. Next, in Section 3, numerical  
95 experiments are employed to demonstrate the proposed method. Then in Section 4, the influence of the  
96 number of pilot points on the results is analyzed. The paper ends with summary and conclusions.

## 97 **2. Methodology**

### 98 *2.1. Algorithm*

99 The ensemble pattern-search algorithm, an extension of the direct sampling (DS) method (a training  
100 image based MPS method) [5], was first proposed by Zhou et al. [47]. The algorithm extended the direct  
101 sampling method in two specific aspects: (1) the pattern is composed of not only the geologic parameters  
102 (e.g, conductivity) but also includes state variables (e.g., piezometric head); (2) an ensemble of realizations  
103 both for the geologic parameters as well as for state variables are used as multiple training images. Li et al.  
104 [48] further improved the algorithm by simultaneously estimating both geologic and state variables, resulting  
105 in a better characterization of permeability/conductivity at both large and small scales.

106 In this work, the ensemble pattern-search algorithm is coupled with the pilot-point concept commonly  
107 used in traditional inverse methods such as sequential self-calibration in order to improve computational  
108 efficiency. Figure 1 shows a flowchart of the EnPAT algorithm. As we can see, if all the gridblocks are pilot  
109 points, the algorithm is completely equivalent to the original implementation proposed by Zhou et al. [47];  
110 on the contrary, if there are no pilot points, the algorithm would become the ensemble direct sampling MPS  
111 method as shown in the paper by Li et al. [48]. In this paper, we employ a finite set of pilot points in order to  
112 render the process of simultaneously searching for both the conductivity and head pattern computationally  
113 fast. Thus, the basic idea is to use the original pattern search method to simulate the conductivity at the  
114 pilot points conditioned to the pattern of both conductivity and head data in the neighborhood of the pilot  
115 point. Subsequently, the ensemble direct sampling MPS (Zhou et al. [47]) is used to complete the subsurface  
116 models conditioned on the previously simulated pilot point values. To achieve this, once the number and  
117 location of pilot points are defined, the corresponding constraints (both  $k$  and  $h$ ) on the nodes are enforced  
118 to simulate the conductivities. When the conductivities at all the pilot points are estimated, the simulation  
119 continues to simulate conductivity values at all remaining nodes conditioned to the values at the pilot point  
120 locations. The pilot point locations are sampled at random from the entire simulation domain.

121 The specific implementation of the EnPAT algorithm can be summarized as follows:

- 122 1. Initialization step. Generate an ensemble of  $N_r$  prior geologic models conditioned to measured con-  
123 ductivity data using traditional MPS methods such as SNESIM [4]. It is assumed that the model is

124 composed of  $N_n$  grid blocks. Here, the initial ensemble can be generated either using a single training  
 125 image (a conceptual model of the aquifer [35]), or using several training images representing the prior  
 126 uncertainty in the geological model.

- 127 2. Forecast step. For each conductivity model (realization), the transient groundwater flow model is  
 128 solved from time zero to  $t$  using a standard numerical simulator, i.e.,

$$\mathbf{Y}_k = f(\mathbf{X}_{k-1}, \mathbf{Y}_0) \quad (1)$$

129  $\mathbf{Y}_k$  and  $\mathbf{Y}_0$  denote the simulated piezometric head at time  $t = k$  and the initial head at time  $t = 0$ ,  
 130 respectively.  $\mathbf{X}_{k-1}$  corresponds to the updated conductivity at time  $t = k - 1$ . The groundwater flow  
 131 model, boundary conditions, sources and other unknown parameters are represented by the transfer  
 132 function  $f$ .

133 The ensemble of conductivity realizations at  $t = k - 1$  (i.e.  $\mathbf{X}_{k-1}$ ) plus the corresponding ensemble  
 134 of simulated piezometric head at  $t = k$  (i.e.  $\mathbf{Y}_k$ ) can be used as the ensemble of training images to  
 135 simulate the ensemble conductivity at  $t = k$  (i.e.  $\mathbf{X}_k$ ) using the following pattern search method.

- 136 3. Ensemble pattern search step. Update the ensemble conductivity  $\mathbf{X}_{k-1}$  based on the observed head  
 137 data at time  $t = k$ .

- 138 (a) Start the loop to estimate the conductivity realization  $r = 1$  at  $t = k$  (i.e.,  $\mathbf{X}_k$ ).
- 139 (b) Generate a random path  $P_i(x), i = 1, 2, \dots, N_n$  visiting all the nodes of the realization, in which  
 140 the first  $N_p$  nodes are used as pilot points. These pilot points are located randomly within the  
 141 domain.
- 142 (c) Start the loop to estimate the conductivity  $X_k^i$  and piezometric head  $Y_k^i$ .
- 143 i. If the conductivity and head data both are known at node  $i$  by measurement (e.g., hard data),  
 144 go to step (d).
- 145 ii. Build the conditioning pattern, which is composed of geologic parameter (e.g., conductivity)  
 146 and state (e.g., head) variables (see Figure 2A). The conductivity and head data in the  
 147 pattern can be either the measured data or the previous estimated data (as in traditional  
 148 sequential simulation). The size of pattern is determined by the predefined maximum number  
 149 of conductivity and head data and the search radius. When the conditioning data are sparse,  
 150 the size of pattern is usually large, which is good for the estimation of conductivity at the  
 151 large scale such as the channels (see Figure 2A); when the available conditioning data becomes  
 152 dense (at later stages of the simulation), the pattern in a smaller area around the simulation



node will be searched, which is beneficial for the simulation of smaller scale features (Figure 2B). This scheme with flexible pattern size is thus equivalent to the multiple grid approach in the sequential Gaussian simulation algorithm proposed by Gómez-Hernández and Journel [3], in terms of reducing the artifacts in the simulation. Also, in order to reduce the computational cost, a spiral search strategy centered on the estimated location  $i$  is considered [2].

Note that, if the simulation is at the early stages and no hard data are available, the conditioning pattern will be composed of only state variables (Figure 2E).

iii. Generate a random path  $P_j, j = 1, 2, \dots, N_r$  visiting all the realizations. Search for a match to the conditional pattern (see Figure 3). Start the loop with a randomly selected realization  $j$ .

A. Search the candidate pattern on the realization  $j$ . The search would be enforced at the node  $i$  because the state (head) is affected by the presence of boundary conditions and source terms. However, this would require a large ensemble in order to find a matching pattern and is consequently computationally expensive. In order to alleviate the computational expense, the 8 nodes (26 nodes in three dimensions) surrounding the center grid block  $i$  are also searched for a matching pattern. Despite using a small area surrounding the node  $i$  to search for the matching pattern in the ensemble, the method has the same accuracy as the basic DS algorithm that uses a single training image (Figure 2C and D) to simulate the pattern of conductivity. At the pilot point location, conductivity is simulated constrained to both conductivity and piezometric head. At other locations, the simulation is conducted with the pattern of only conductivity; in other words, the ensemble DS is used to extrapolate the conductivity on the basis of the previously simulated pilot point.

B. Calculate the distances,  $d_j^X$  for conductivity and  $d_j^Y$  for head, between the conditional pattern and the candidate pattern found during the search. Specific distance functions will be described in section 2.2.

C. Compare the calculated distance values with the predefined tolerance values. Specifically:

$$\left\{ \begin{array}{ll} \text{Case 1 : } X_k^i = X_{k-1}^j, Y_k^i = Y_k^j, & \text{if } d_j^X < \xi^X \text{ and } d_j^Y < \xi^Y . \\ \text{Case 2 : } X_k^i = X_{k-1}^j, & \text{if } d_j^X < \xi^X . \\ \text{Case 3 : } Y_k^i = Y_k^j, & \text{if } d_j^Y < \xi^Y . \end{array} \right. \quad (2)$$

180 where  $\xi^X$  and  $\xi^Y$  are the predefined tolerance values for the conductivity and head,  
 181 respectively. Case 1 denotes simulation at the pilot point locations (Figure 2A and B), in  
 182 which case tolerances are applied to both the conductivity and piezometric head patterns  
 183 in order to find the match. Case 2 corresponds to simulation of the non-pilot point  
 184 locations (Figure 2C and D), in which case a tolerance is applied only to the conductivity  
 185 pattern in order to find the closest match. Case 3 occurs when hard data are not available  
 186 and only the head data are used to construct the pattern (Figure 2E and F). In that case  
 187 a tolerance is only applied to the pattern of piezometric head in order to find the closest  
 188 match.

189 If the computed distance is less than the specified tolerance, go to step (d).

190 D. Otherwise, set  $j = j + 1$ , and go to step A.

- 191 iv. If at the end of previous steps the tolerance criteria are not met, then the pattern with  
 192 the smallest distances is retained and the corresponding grid block ( $n$ ) conductivity value is  
 193 assigned as the simulated one, i.e.,

$$\begin{cases} \text{Case 1 : } X_k^i = X_{k-1}^n, Y_k^i = Y_k^n, & \text{if } d_{min}^X = d_n^X \text{ and } d_{min}^Y = d_n^Y . \\ \text{Case 2 : } X_k^i = X_{k-1}^n, & \text{if } d_{min}^X = d_n^X . \\ \text{Case 3 : } Y_k^i = Y_k^n, & \text{if } d_{min}^Y = d_n^Y . \end{cases} \quad (3)$$

194 where  $d_{min}^X$  and  $d_{min}^Y$  are the smallest distances for the conductivity and piezometric head,  
 195 respectively. The simulated values will be used as the conditioning data for the subsequent  
 196 simulation.

197 (d) Set  $i = i + 1$  and go to step (c) until all nodes are estimated.

198 (e) Set  $r = r + 1$  and go to step (a) until all realizations are simulated.

- 199 4. Loop back. The forecast and pattern search loop is applied again on the updated conductivity field for  
 200 the next time step of observations (i.e., set  $k = k + 1$  and go to step 2 until all the time step observation  
 201 data are conditioned).

202 Even though both conductivity and head are simultaneously simulated using the pattern-search approach  
 203 in order to capture the variability of the aquifer attributes at multiple scales [48], only the simulated con-  
 204 ductivity is kept for updating at the next time step. This is because the updated piezometric heads might  
 205 not be consistent with those simulated by the full-physical simulator because of complexity of well/boundary

206 conditions, etc. The training images for piezometric head for pattern search at all subsequent time steps are  
 207 obtained by re-running the forward simulator from time step zero. This does add to the computational cost,  
 208 but it is similar to the implementation in EnKF-based methods such as static EnKF and re-start EnKF [e.g.,  
 209 54, 44, 57, 45].

210 Unlike in previous implementations [47, 48] where the pattern search constrained to both conductivity  
 211 and head is performed at all nodes in the domain, in this new algorithm, pattern search is performed only at  
 212 predefined pilot points. At all other non-pilot point locations, the pattern is composed of only conductivity  
 213 and the simulation is conducted by pattern search through the criteria presented in Case 2 above. The  
 214 implementation of the full pattern search at pilot point locations not only accelerates the simulation, but  
 215 also improves the quality of updated conductivity field in terms of preserving the large scale connectivity.  
 216 The pilot points are randomly located within the field, and they are changed from one realization to the  
 217 next as well as from one time step to the next.

## 218 2.2. Distance functions

219 In the algorithm, distance functions play a key role to quantify the candidate pattern. Follow the same  
 220 approach as in Zhou et al. [47]. The Manhattan distance is computed for categorical variables (i.e., rock  
 221 facies or homogeneous conductivities within facies) and the weighted Euclidian distance function is computed  
 222 for continuous variables (for example, piezometric head or continuous conductivities).

- 223 • Manhattan distance for the categorical variables:

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

$$a_i = \begin{cases} 0, & \text{if } d(x_i) = p(x_i) \\ 1, & \text{otherwise} \end{cases} \quad (4)$$

224 where  $n$  is the numbers of nodes in the pattern,  $d(x_n)$  is the conditioning pattern,  $p(x_n)$  is the matching  
 225 pattern,  $x_i$  represents the members of the pattern.

- 226 • Weighted Euclidean distance for the continuous variables

$$d\{d(x_n), p(x_n)\} = \left( \frac{1}{\sum_{i=1}^n h_i^{-1}} \sum_{i=1}^n h_i^{-1} \frac{|d(x_i) - p(x_i)|^2}{d_{max}^2} \right)^{1/2} \quad d \in [0, 1] \quad (5)$$

227 where  $h_i$  is the Euclidean distance between node  $i$  and the node being simulated, therefore giving more  
228 weight to dissimilarities of the closest nodes,  $d_{max}$  is the maximum absolute difference  $|d(x_i) - p(x_i)|$ .

### 229 3. Synthetic example

#### 230 3.1. Reference field

231 A single-phase transient groundwater flow problem is set up to demonstrate the effectiveness of the  
232 proposed method. The aquifer is discretized into  $50 \times 50 \times 1$  grid blocks, with the block dimensions of 1  
233  $m \times 1 m \times 1 m$ . The training image from Strebelle [4] is used as the conceptual geologic model with high  
234 conductivity sand and low conductivity shale (see Figure 4A). The SNESIM code developed by Strebelle [4]  
235 is employed to generate the reference facies map, in which the proportion of sand is set as 30%. Constant  
236 hydraulic conductivity values with 10 and  $10^{-4} m/d$  are assigned to sand and shale, respectively. The  
237 variance of log-conductivities is about 5.3. The reference model is shown in Figure 4B.

238 The aquifer is assumed to be confined and modeled with constant head boundaries at  $x = 0 m$  and  $x = 50$   
239  $m$  (i.e.,  $h = 0 m$  at those boundaries) and with no flow boundaries at the remaining model faces (see Figure  
240 4B). The initial head is assumed zero over the entire domain. Specific storage is set as constant and equal  
241 to  $0.01 m^{-1}$ . A pumping well (well #2) located at the center of the domain extracts water at a constant  
242 flow rate of  $25 m^3/d$ . There are three observation wells located within the domain. The spatial locations  
243 of the 4 wells are shown in Figure 4B. The simulation time is 30 *days* and is discretized into 10 time steps  
244 following a geometric sequence of ratio 1.2.

245 The finite-difference code MODFLOW [58] is used to solve the transient flow equation (1). The piezo-  
246 metric head data for the first five time steps (8.6 *days*), collected from the 4 wells are used as conditioning  
247 data to update the initial model parameters. No measurement error is considered as well as no measured  
248 conductivity data in this example. The measurement error could be implicitly accounted for by increasing  
249 the distance tolerance value for each variable.

250 Five hundred unconditional facies realizations are generated using SNESIM and then populated with  
251 constant conductivity values. The parameters used in the EnPAT are listed in Table 1. Note that: (1) the  
252 tolerance values both for the conductivity and head are defined as zero so that the best simulated value  
253 (corresponding to the minimum distance) is achieved at the pilot point locations, even though this increases  
254 the computational cost; (2) The number of pilot point is set at 500. Sensitivity of the results to the number  
255 of pilot point locations will be studied in section 4. The sensitivity to other parameters such as tolerance  
256 and number of conditioning data in the pattern is referred to the work by Meerschman et al. [59].

257 *3.2. Evaluation criteria*

258 A set of metrics are computed in order to quantify the performance of proposed method.

- 259 1. Check the ensemble mean and variance of conductivity maps, that can be used to check the reproduction  
260 of the main patterns observed in the reference and to display the spatial uncertainty depicted by the  
261 updated ensemble of conductivity realizations.
- 262 2. Calculate the average absolute error (*AAE*) and average ensemble spread (*AES*), that can be used to  
263 quantify the accuracy and uncertainty of simulated results. They are defined as follows:

$$\begin{aligned} \text{AAE}(x) &= \frac{1}{N_n} \sum_{i=1}^{N_n} \frac{1}{N_r} \sum_{r=1}^{N_r} |x_{i,r} - x_{\text{ref},i}| \\ \text{AES}(x) &= \frac{1}{N_n} \sum_{i=1}^N \sigma_{x_i}^2 \end{aligned} \tag{6}$$

264 where  $x_{i,r}$  is conductivity at location  $i$  for realization  $r$ ,  $x_{\text{ref},i}$  is the reference conductivity value at  
265 location  $i$ ,  $N_n$  is the number of nodes,  $N_r$  is the number of realizations in the ensemble and  $\sigma_{x_i}^2$  is the  
266 ensemble variance at location  $i$ .

- 267 3. Calculate the ensemble connectivity of conductivity. Here, the code CONNEC3D developed by Pardo-  
268 Igúzquiza and Dowd [60] is used to compute the connectivity. In this approach, the conductivity values  
269 are first converted to indicators and then the connectivity of indicators can be calculated, dependent on  
270 the user defined spatial template or connectivity direction. The definition and computation of connec-  
271 tivity is referred the study by Pardo-Igúzquiza and Dowd [60]. In our case, horizontal connectivity (i.e.,  
272 east-west direction) is calculated because the reference conductivity field shows dominant connectivity  
273 in that direction.

274 *3.3. Results*

275 *3.3.1. Hydraulic conductivity characterization*

276 Using the EnPAT data integration algorithm, an ensemble of 500 conductivity realizations are obtained  
277 for each conditioning time step. Here, all the evaluation is based on the indicator transform of conductivity,  
278 i.e., high-conductivity has an indicator value of 1 and low-conductivity shale of 0. Figure 5 shows the  
279 ensemble mean and variance of conductivity at three time steps. Figure 6 displays two randomly selected  
280 realizations before and after conditioning to the dynamic data. Figure 7 shows the spatial distribution of  
281 pilot points and the simulated values at those locations in two realizations. The evolution of the calculated

282  $AAE$  and  $AES$  with time is shown in Figure 8. Figure 9 displays the connectivity calculated for all members  
283 of the ensemble as well as the ensemble average before and after data conditioning.

284 Analyzing the evolution of ensemble mean and variance maps with time, it is evident that (a) without any  
285 conditioning data, the mean and variance of the initial ensemble is spatially constant because the channels  
286 are randomly distributed in space; (b) the main pattern of channels gradually converges to the reference,  
287 and also the uncertainty of ensemble conductivity is reduced. From a quantitative point of view, Figure 8  
288 also shows that the  $AAE$  is reduced by about 30% and  $AES$  by about 60% after conditioning to all available  
289 head data. Both  $AAE$  and  $AES$  attain a plateau after time step 4 implying that further conditioning to  
290 head data will not improve the characterization of the conductivity field. Other information such as solute  
291 concentration or a denser network of wells would be needed to further resolve the conductivity field. The  
292 connectivity statistics in Figure 9 also demonstrate an extreme reduction in uncertainty after conditioning  
293 to head data. More specifically, the ensemble mean of connectivity is close to the reference and the spread of  
294 connectivity over the ensemble realizations is reduced after conditioning to the head data. The two individual  
295 realizations in Figure 6 depict the channels before and after data conditioning. We can conclude that the  
296 channels in the initial model deviate considerably from the reference while the updated model exhibits similar  
297 spatial distribution of channels as in the reference. Figure 7 shows the pilot point locations used for two of  
298 the realizations. Judging by the conductivity values at the pilot point locations, we can conclude that the  
299 spatial distribution of channels has already been broadly represented in the pilot point values. This further  
300 supports our implementation of conditioning to head and conductivity patterns at pilot point locations and  
301 subsequent spatial interpolation of only conductivity values away from pilot point locations.

302 Based on the results above it can be conjectured that the EnPAT algorithm can be efficiently used to  
303 integrate head data into the simulation of conductivity fields. Note that the results further indicate that in  
304 the absence of hard conductivity data here, features such as channels can be simulated based on head data  
305 and based on learning the relationship between conductivity and the dynamic response.

### 306 3.3.2. Flow predictions

307 The ultimate objective of the dynamic data integration process is to make better prediction of future  
308 performance of the aquifer flow process. The ensemble of geologic models were rerun from time zero using  
309 the updated conductivity after conditioning to all the head data.

310 Figure 10 shows the variation in predicted head values over the ensemble and the corresponding average  
311 over the ensemble for the 4 wells. The head data for the first 8.6 *days* was used for conditioning and the data  
312 for the remaining period is used to assess the prediction accuracy of the models. We can see that without

313 any conditioning, the spread in head values is large and the average deviates significantly from the reference.  
 314 After conditioning to the head data, the ensemble average of predicted head is close to the reference and the  
 315 variation in predicted values (uncertainty) is also reduced.

#### 316 4. Sensitivity to the number of pilot points

317 One key issue with this modified EnPAT algorithm is the optimization of the number and location of the  
 318 pilot points. Figure 11 displays the evolution of the ensemble mean of updated conductivity values for the  
 319 first time step using different number of pilot points within EnPAT (i.e.,  $Np = 20, 100, 500$  and  $1500$ ). We  
 320 can see that the pattern exhibited by the ensemble mean is best reproduced with 500 or more pilot points.  
 321 Clearly, if there are no pilot points then no conditioning to head data is achieved and in the absence of any  
 322 hard conductivity data, the ensemble mean map would be constant as shown in Figure 5 ( $nT = 0$ ). On  
 323 the other hand, if there are a larger number of pilot points, the computational expense is high and also the  
 324 simulated images show artifacts due to the strong constraints enforced during the pattern search process  
 325 (see Figure 4 of paper by Zhou et al. [47] taking all the gridblocks as pilot points). The speed up ( $S_p$ ) has  
 326 been calculated to evaluate the computational gain, i.e.,

$$S_p = \frac{T}{T_p} \quad (7)$$

327 where  $T$  is the CPU time consumed under the original EnPAT implementation (i.e., all the nodes are pilot  
 328 points and needed to be simulated with constraints of both conductivity and piezometric head),  $T_p$  is the  
 329 CPU time using  $p$  pilot points. Figure 12 shows the speedup with the number of pilot points. As expected,  
 330 speed up is reduced as more pilot points are used.

331 Choosing a reasonable number of pilot points not only reduces the computational cost, but also improves  
 332 the quality of the updated conductivity fields. In the current implementation, no further postprocess pro-  
 333 cedure is required while in the original EnPAT algorithm proposed by Zhou et al. [47], the postprocess is  
 334 needed to remove the artifacts which are not exhibited in the training images. In this example, using about  
 335 one fifth of the total number of nodes, the speed up is about 2.5 times, and the main spatial pattern of  
 336 the channels is exhibited in the ensemble mean map, similar to the original implementation of EnPAT (see  
 337 Figure 11 and 12). A practical guideline for selecting the number of pilot points could be such that the  
 338 average distance between pilot points is smaller than the channel width in order to capture the details of  
 339 the curvilinear structure.

## 340 5. Discussion

341 The EnPAT algorithm shares some features with the EnKF [21]: (1) an ensemble of realizations is used,  
342 and thus uncertainty in flow and transport predictions can be assessed; (2) a full-physical simulator is inte-  
343 grated as a black-box in both methods; (3) the correlation between parameter and state is explicitly estimated  
344 through the ensemble realizations. However, a significant difference exists in the way the parameters are up-  
345 dated. In the EnKF, the parameter is updated as the initial value plus the Kalman Gain weighted dynamic  
346 data mismatch. That is, the complex non-linear relationship between the model parameters and the dynamic  
347 response is approximated by the Kalman Gain term that is computed based on the covariance, and thus the  
348 updating procedure is optimal only if the parameter and state variables are jointly multiGaussian. In the  
349 EnPAT algorithm, on the other hand, the model parameters are simulated using the pattern-search method,  
350 in which the correlation between parameter and the state variables (dynamic response) is described in terms  
351 of a spatial pattern (i.e. beyond two-point covariance), and thus the parameter and state do not necessarily  
352 follow a multiGaussian distribution. Furthermore the curvilinear/complex structure of the aquifer parameter  
353 is explicitly preserved through the multipoint statistics.

354 In this paper, we used a synthetic 2D example to demonstrate the robustness and effectiveness of proposed  
355 method. In real case, the aquifer could be 3D in nature and the simulation of channeled aquifer will be more  
356 CPU demanding and time consuming in 3D than in 2D but, in principle, the method is general in its  
357 formulation and is applicable to 3D domains. It is however possible that the flow response in 3D systems  
358 are not as seriously impacted by complex reservoir heterogeneity which in turn can impact the quality and  
359 variability of the inverted permeability fields [61]. This issue will be investigated in a subsequent paper.

## 360 6. Conclusion

361 Inverse modeling is helpful for modeling and prediction of subsurface flow and transport processes. Most  
362 of the current approaches are based on an optimization framework. Recently, the EnPAT algorithm [47, 48]  
363 was developed to condition non-Gaussian conductivity fields to dynamic data. Unlike current approaches,  
364 the EnPAT is based on pattern search and simulation and does not explicitly involve minimization of an  
365 objective function. In this approach, the joint pattern of both conductivity and piezometric head data is  
366 considered, and the direct sampling multiple-point statistics method [5] is applied to find the matching  
367 pattern from an ensemble of aquifer models. Consequently, the updated conductivity realizations not only  
368 are conditioned to the available piezometric head data, but also preserve complex curvilinear structure as  
369 described by the spatial pattern.



370 In this paper, we integrate the pilot point concept [16, 19] into the EnPAT method. Specifically, the  
371 conductivity at the pilot point location is simulated by searching for the pattern composed of both conduc-  
372 tivity and piezometric head. Subsequently, the conductivity at the remaining locations is simulated using  
373 the original Direct Sampling algorithm. The main difference being that the matching pattern is searched  
374 over an ensemble of training images instead of a single training image.

375 We demonstrate the proposed method using a synthetic 2D transient flow example. Our results indicate  
376 that: (1) the computational cost of EnPAT is reduced by using pilot points while at the same time the  
377 quality of updated conductivity realizations is improved; (2) indirect data such as piezometric head have  
378 information that helps us identify the complex spatial characteristics of the conductivity field even when  
379 measured conductivity data may not be available.

380 **Acknowledgements.** The first three authors gratefully acknowledge the financial support by DOE  
381 through project DE-FE0004962. The fourth author acknowledges the financial support by the Spanish  
382 Ministry of Science and Innovation through project CGL2011-23295. The authors also wish to thank  
383 Wolfgang Nowak as well as two anonymous reviewers for their comments, which helped improving the final  
384 version of the manuscript.

## 385 References

- 386 [1] Journel A. Geostatistics for conditional simulation of ore bodies. *Economic Geology* 1974;69(5):673–87.
- 387 [2] Deutsch CV, Journel AG. *GSLIB, Geostatistical Software Library and User’s Guide*. New York: Oxford  
388 University Press; 1992.
- 389 [3] Gómez-Hernández JJ, Journel AG. Joint sequential simulation of multigaussian fields. In: *Geostatistics*  
390 *Troia92*. Springer; 1993, p. 85–94.
- 391 [4] Strebelle S. Conditional simulation of complex geological structures using multiple-point statistics.  
392 *Mathematical Geology* 2002;34(1):1–21.
- 393 [5] Mariethoz G, Renard P, Straubhaar J. The direct sampling method to perform multiple-point geosta-  
394 tistical simulaitons. *Water Resources Research* 2010;46(11).
- 395 [6] Huang Y, Srinivasan S. Efficient conditional simulation of spatial patterns using a pattern-growth  
396 algorithm. *Geostatistics Oslo 2012*;2012::209–20.

- 397 [7] Yoram R, et al. Applied stochastic hydrogeology. Oxford University Press; 2003.
- 398 [8] Dagan G, Neuman SP. Subsurface flow and transport: a stochastic approach. Cambridge University  
399 Press; 2005.
- 400 [9] Gómez-Hernández JJ. Complexity. Ground Water 2006;44(6):782–5.
- 401 [10] Sun N. Inverse problems in groundwater modeling. Kluwer Academic Dordrecht; 1994.
- 402 [11] Zimmerman D, De Marsily G, Gotway C, Marietta M, Axness C, Beauheim R, et al. A comparison  
403 of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective  
404 transport by groundwater flow. Water Resources Research 1998;34(6):1373–413.
- 405 [12] Carrera J, Alcolea A, Medina A, Hidalgo J, Slooten L. Inverse problem in hydrogeology. Hydrogeology  
406 Journal 2005;13(1):206–22.
- 407 [13] Hendricks Franssen H, Alcolea A, Riva M, Bakr M, van der Wiel N, Stauffer F, et al. A comparison of  
408 seven methods for the inverse modelling of groundwater flow. application to the characterisation of well  
409 catchments. Advances in Water Resources 2009;32(6):851–72.
- 410 [14] Oliver D, Chen Y. Recent progress on reservoir history matching: a review. Computational Geosciences  
411 2011;15(1):185–221.
- 412 [15] Zhou H, Gómez-Hernández J, Li L. Inverse methods in hydrogeology: evolution and recent trends.  
413 Advances in Water Resources 2013;Under review.
- 414 [16] De Marsily G, Lavedan G, Boucher M, Fasanino G. Interpretation of interference tests in a well field  
415 using geostatistical techniques to fit the permeability distribution in a reservoir model. Geostatistics  
416 for Natural Resources Characterization, Part 1984;2:831–49.
- 417 [17] Carrera J, Neuman S. Estimation of aquifer parameters under transient and steady state condi-  
418 tions: 1. Maximum likelihood method incorporating prior information. Water Resources Research  
419 1986;22(2):199–210.
- 420 [18] Kitanidis P. Quasi-linear geostatistical theory for inversing. Water Resources Research  
421 1995;31(10):2411–9.
- 422 [19] Gómez-Hernández JJ, Sahuquillo A, Capilla JE. Stochastic simulation of transmissivity fields conditional  
423 to both transmissivity and piezometric data, 1, Theory. Journal of Hydrology 1997;203(1–4):162–74.

- 424 [20] Hu LY. Gradual deformation and iterative calibration of gaussian-related stochastic models. *Math*  
425 *Geology* 2000;32(1):87–108.
- 426 [21] Evensen G. The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean*  
427 *dynamics* 2003;53(4):343–67.
- 428 [22] Capilla J, Jaime Gómez-Hernández J, Sahuquillo A. Stochastic simulation of transmissivity fields con-  
429 ditional to both transmissivity and piezometric data 2. demonstration on a synthetic aquifer. *Journal*  
430 *of Hydrology* 1997;203(1):175–88.
- 431 [23] Capilla J, Jaime Gómez-Hernández J, Sahuquillo A. Stochastic simulation of transmissivity fields con-  
432 ditional to both transmissivity and piezometric head data 3. application to the culebra formation at the  
433 waste isolation pilot plan (wipp), New Mexico, USA. *Journal of Hydrology* 1998;207(3):254–69.
- 434 [24] Hendricks Franssen H, Gómez-Hernández J, Sahuquillo A. Coupled inverse modelling of groundwater  
435 flow and mass transport and the worth of concentration data. *Journal of Hydrology* 2003;281(4):281–95.
- 436 [25] LaVenue A, Pickens J. Application of a coupled adjoint sensitivity and kriging approach to calibrate a  
437 groundwater flow model. *Water Resources Research* 1992;28(6):1543–69.
- 438 [26] Wen X, Deutsch C, Cullick A. High-resolution reservoir models integrating multiple-well production  
439 data. *SPE Journal* 1998;3(4):344–55.
- 440 [27] Wen X, Lee S, Yu T. Simultaneous integration of pressure, water cut, 1 and 4-d seismic data in  
441 geostatistical reservoir modeling. *Mathematical geology* 2006;38(3):301–25.
- 442 [28] Wen X, Capilla J, Deutsch C, Gómez-Hernández J, Cullick A. A program to create permeability fields  
443 that honor single-phase flow rate and pressure data. *Computers & Geosciences* 1999;25(3):217–30.
- 444 [29] Heidari L, Gervais V, Ravalec ML, Wackernagel H. History matching of petroleum reservoir models by  
445 the ensemble kalman filter and parameterization methods. *Computers & Geosciences* 2013;55:84–95.
- 446 [30] Gómez-Hernández J, Wen X. To be or not to be multi-Gaussian? a reflection on stochastic hydrogeology.  
447 *Advances in Water Resources* 1998;21(1):47–61.
- 448 [31] Wen X, Gómez-Hernández J. Numerical modeling of macrodispersion in heterogeneous media: a com-  
449 parison of multi-gaussian and non-multi-gaussian models. *Journal of Contaminant Hydrology* 1998;30(1-  
450 2):129–56.

- 451 [32] Zinn B, Harvey CF. When good statistical models of aquifer heterogeneity go bad: A comparison of  
452 flow, dispersion, and mass transfer in connected and multivariate gaussian hydraulic conductivity fields.  
453 Water Resources Research 2003;39(3).
- 454 [33] Kerrou J, Renard P, Hendricks Franssen H, Lunati I. Issues in characterizing heterogeneity and con-  
455 nectivity in non-multigaussian media. Advances in Water Resources 2008;31(1):147–59.
- 456 [34] Renard P, Allard D. Connectivity metrics for subsurface flow and transport. Advances in Water  
457 Resources 2011;51:168–96.
- 458 [35] Guardiano F, Srivastava R. Multivariate geostatistics: beyond bivariate moments. In: Soares A, editor.  
459 Geostatistics-Troia. Dordrecht: Kluwer Academic Publ; 1993, p. 133–44.
- 460 [36] Hu LY, Chugunova T. Multiple-point geostatistics for modeling subsurface heterogeneity: A compre-  
461 hensive review. Water Resources Research 2008;44(11).
- 462 [37] Haslauer C, Guthke P, Bárdossy A, Sudicky E. Effects of non-gaussian copula-based hydraulic conduc-  
463 tivity fields on macrodispersion. Water Resources Research 2012;48(7).
- 464 [38] Caers J. Geostatistical history matching under training-image based geological model constraints. SPE  
465 Annual Technical Conference and Exhibition, 2002;:SPE 77429.
- 466 [39] Alcolea A, Renard P. Blocking Moving Window algorithm: Conditioning multiple-point simulations to  
467 hydrogeological data. Water Resources Research 2010;46.
- 468 [40] Mariethoz G, Renard P, Caers J. Bayesian inverse problem and optimization with iterative spatial  
469 resampling. Water Resources Research 2010;46(11).
- 470 [41] Sun AY, Morris AP, Mohanty S. Sequential updating of multimodal hydrogeologic parameter fields  
471 using localization and clustering techniques. Water Resources Research 2009;45(7).
- 472 [42] Sarma P, Chen W. Generalization of the ensemble kalman filter using kernels for nongaussian random  
473 fields. In: SPE Reservoir Simulation Symposium. 2009,.
- 474 [43] Zhou H, Gómez-Hernández J, Hendricks Franssen H, Li L. An approach to handling non-gaussianity  
475 of parameters and state variables in ensemble kalman filtering. Advances in Water Resources  
476 2011;34(7):844–64.

- 477 [44] Jafarpour B, Khodabakhshi M. A probability conditioning method (PCM) for nonlinear flow data  
478 integration into multipoint statistical facies simulation. *Mathematical Geosciences* 2011;43(2):133–64.
- 479 [45] Hu LY, Zhao Y, Liu Y, Scheepens C, Bouchard A. Updating multipoint simulatings using the ensemble  
480 Kalman filter. *Computers & Geosciences* 2013;51:7–15.
- 481 [46] Lange K, Frydendall J, Cordua KS, Hansen TM, Melnikova Y, Mosegaard K. A frequency matching  
482 method: solving inverse problems by use of geologically realistic prior information. *Mathematical*  
483 *Geosciences* 2012;44(7):783–803.
- 484 [47] Zhou H, Gómez-Hernández J, Li L. A pattern-search-based inverse method. *Water Resources Research*  
485 2012;48(3).
- 486 [48] Li L, Srinivasan S, Zhou H, Gómez-Hernández JJ. Simultaneous estimation of both geologic and reservoir  
487 state variables within an ensemble-based multiple-point statistic framework. *Mathematical Geosciences*  
488 2012;under review.
- 489 [49] Fu J, Gómez-Hernández JJ. A Blocking Markov Chain Monte Carlo Method for Inverse Stochastic  
490 Hydrogeological Modeling. *Mathematical Geosciences* 2009;41(2):105–28.
- 491 [50] Wen XH, Chen W. Real-time reservoir model updating using ensemble Kalman filter. In: *SPE reservoir*  
492 *simulation symposium*. 2005,.
- 493 [51] Gu Y, Oliver D. The ensemble Kalman filter for continuous updating of reservoir simulation models.  
494 *Journal of Energy Resources Technology* 2006;128:79–87.
- 495 [52] Chen Y, Zhang D. Data assimilation for transient flow in geologic formations via ensemble Kalman  
496 filter. *Advances in Water Resources* 2006;29(8):1107–22.
- 497 [53] Aanonsen S, Nævdal G, Oliver D, Reynolds A, Valles B. The ensemble Kalman filter in reservoir  
498 engineering—a review. *Spe Journal* 2009;14(3):393–412.
- 499 [54] Nowak W. Best unbiased ensemble linearization and the quasi-linear Kalman ensemble generator. *Water*  
500 *Resources Research* 2009;45(4).
- 501 [55] Li L, Zhou H, Hendricks Franssen HJ, Gómez-Hernández JJ. Modeling transient groundwater flow by  
502 coupling ensemble Kalman filtering and upscaling. *Water Resources Research* 2012;48(1).

- 503 [56] Zhou H, Li L, Franssen HJH, Gómez-Hernández JJ. Pattern recognition in a bimodal aquifer using the  
504 normal-score ensemble kalman filter. *Mathematical Geosciences* 2012;44(2):169–85.
- 505 [57] Schöniger A, Nowak W, Hendricks Franssen HJ. Parameter estimation by ensemble Kalman filters  
506 with transformed data: Approach and application to hydraulic tomography. *Water Resources Research*  
507 2012;48(4).
- 508 [58] Harbaugh AW, Banta ER, Hill MC, McDonald MG. MODFLOW-2000, the U.S. Geological Survey  
509 modular ground-water model. Reston, VA, Denver, CO: U.S. Geological Survey, Branch of Information  
510 Services; 2000.
- 511 [59] Meerschman E, Pirot G, Mariethoz G, Straubhaar J, Meirvenne M, Renard P. A practical guide to  
512 performing multiple-point statistical simulations with the direct sampling algorithm. *Computers &*  
513 *Geosciences* 2013;52:307–24.
- 514 [60] Pardo-Igúzquiza E, Dowd P. CONNEC3D: a computer program for connectivity analysis of 3D random  
515 set models. *Computers & geosciences* 2003;29(6):775–85.
- 516 [61] Fogg GE. Groundwater flow and sand body interconnectedness in a thick, multiple-aquifer system.  
517 *Water Resources Research* 1986;22(5):679–94.

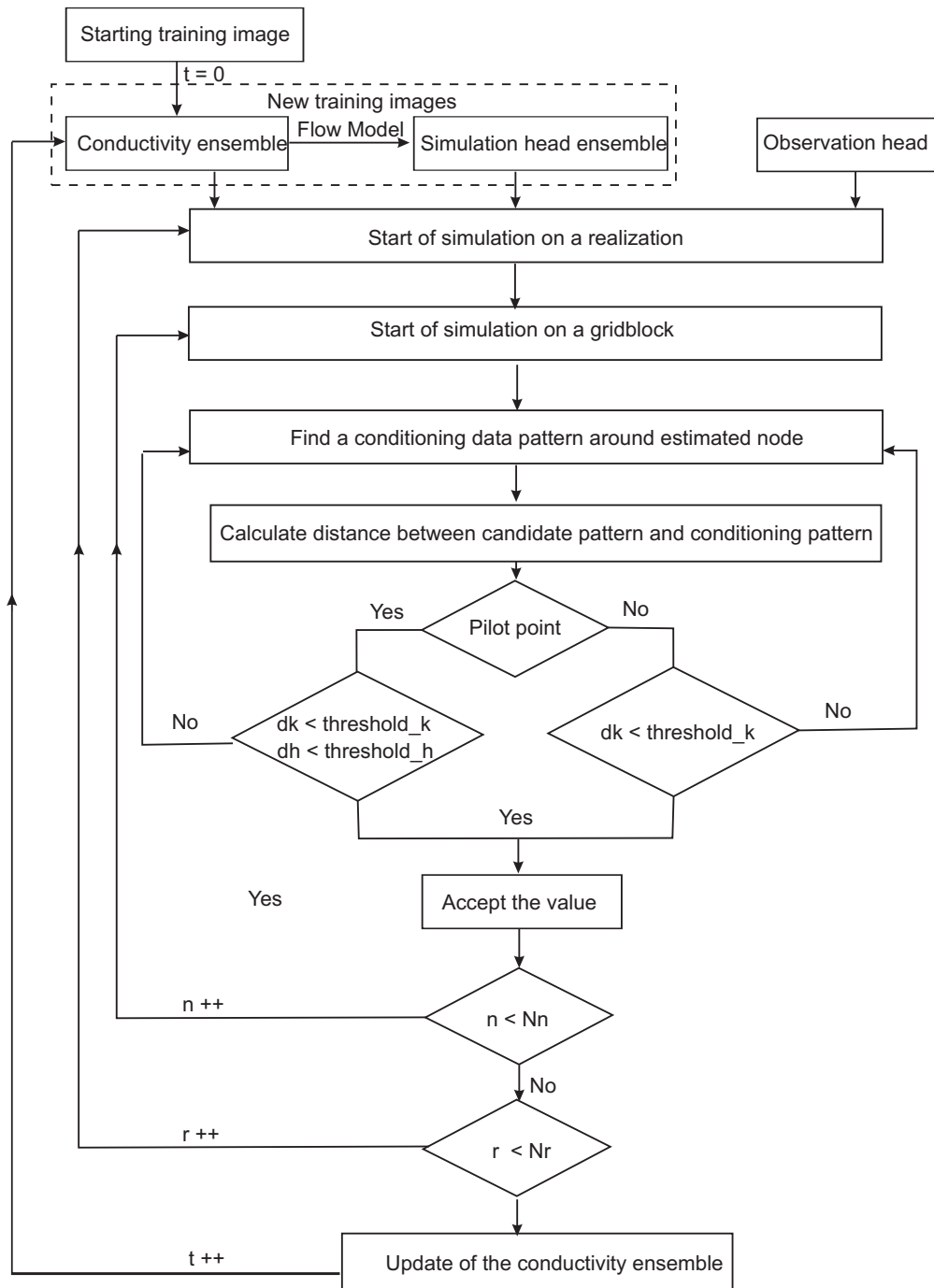


Figure 1: Flowchart of the algorithm.

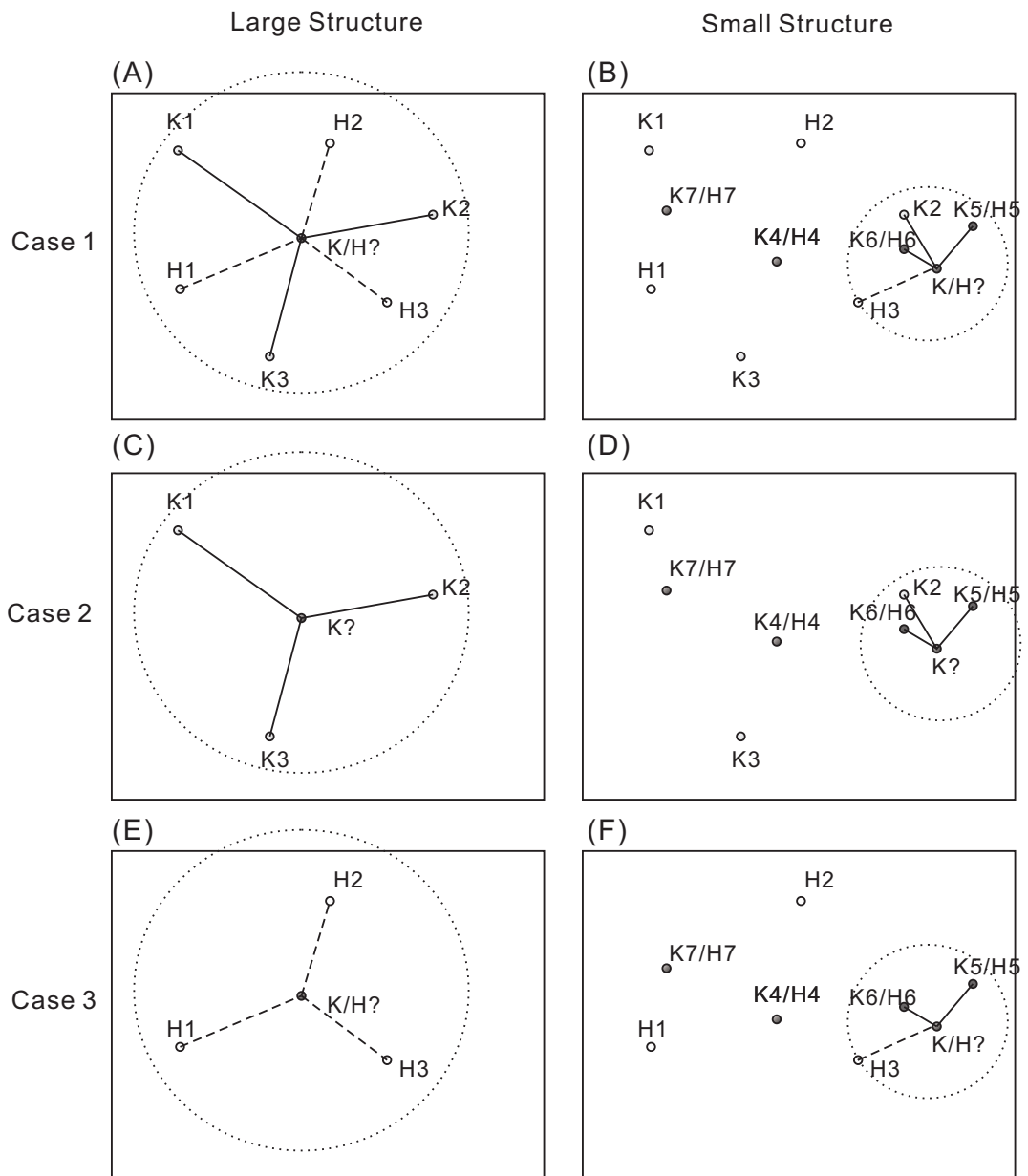


Figure 2: Spatial evolution of the pattern in the EnPAT algorithms. Case 1 denotes the pattern at the pilot points locations; Case 2 is one at the non-pilot point locations and Case 3 corresponds to the pattern in case there is only the measured dynamic data.



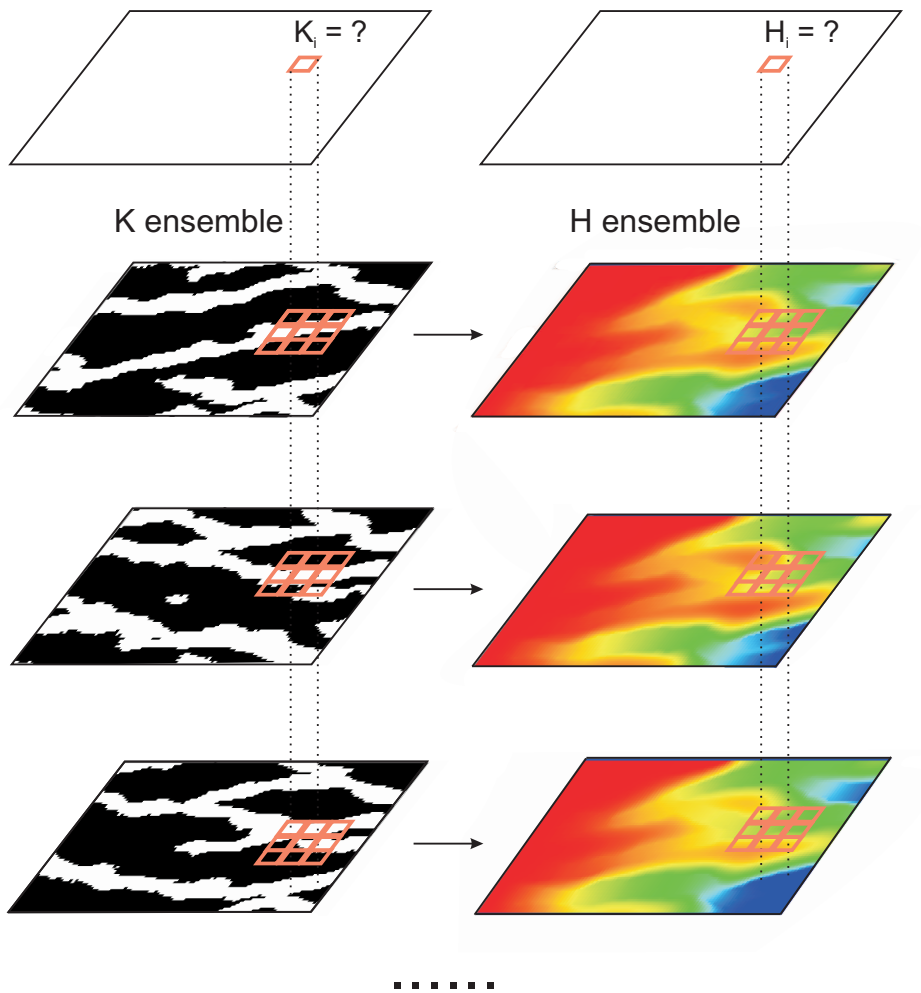


Figure 3: Pattern search with ensemble training images.

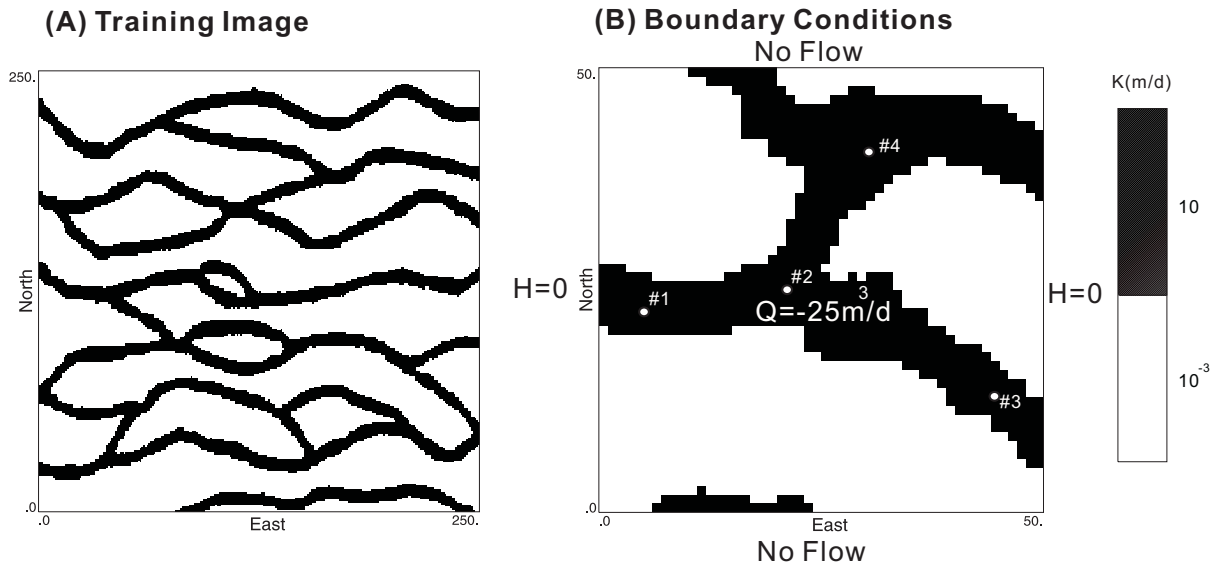


Figure 4: (A) Training image (B) Reference conductivity, boundary conditions of flow model and observation wells.

Table 1: Parameters used in EnPAT.

Search radius for conductivity	25 m
Search radius for head	25 m
Max. number of element in the pattern for conductivity	10
Max. number of element in the pattern for head	10
Distance function for conductivity	Manhattan
Distance function for head	Weighted Euclidian
Distance tolerance for conductivity	0.0
Distance tolerance for head	0.0
Number of pilot point	500

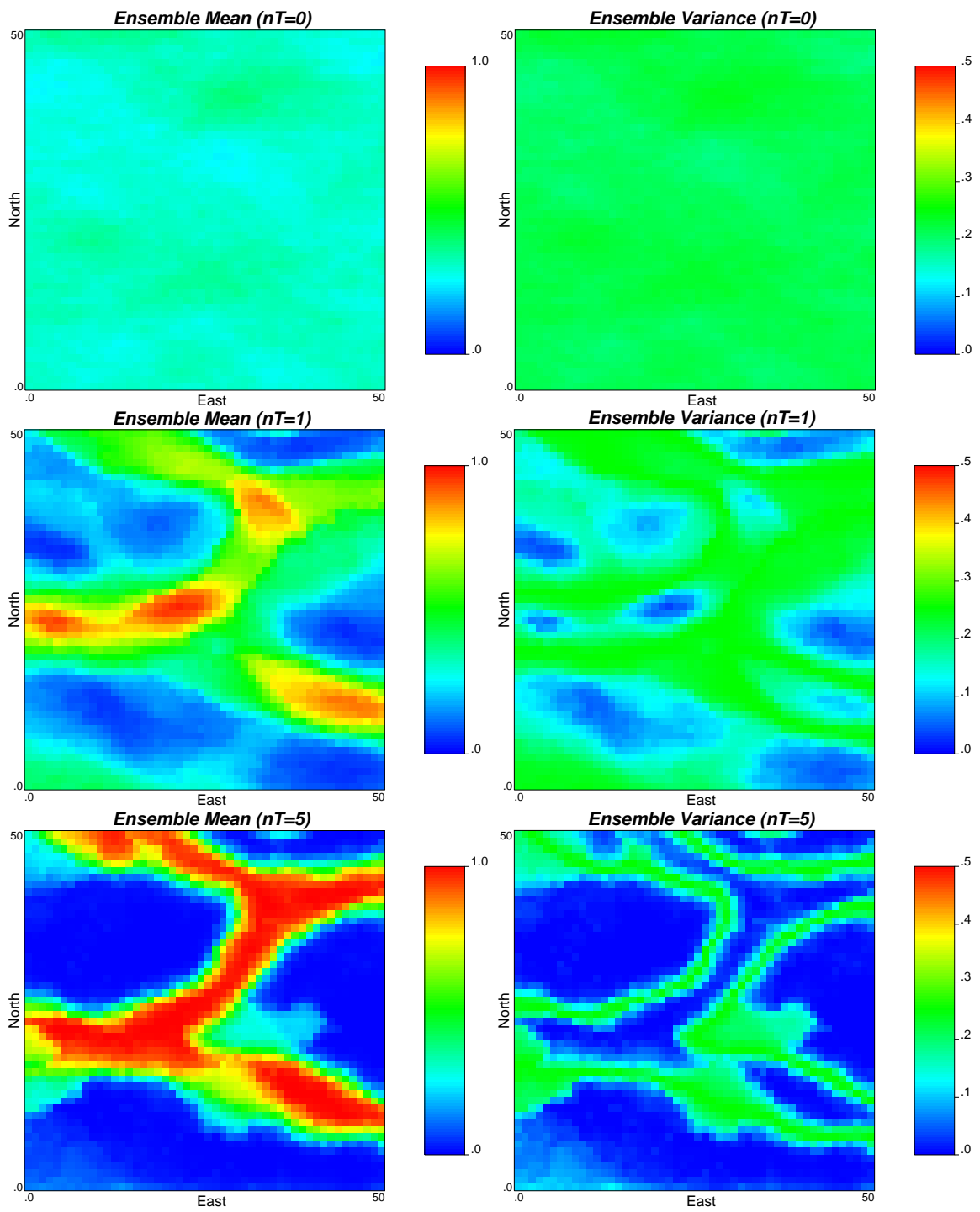


Figure 5: The ensemble mean and variance of conductivity indicator values (i.e., shale with 0 and sand with 1)

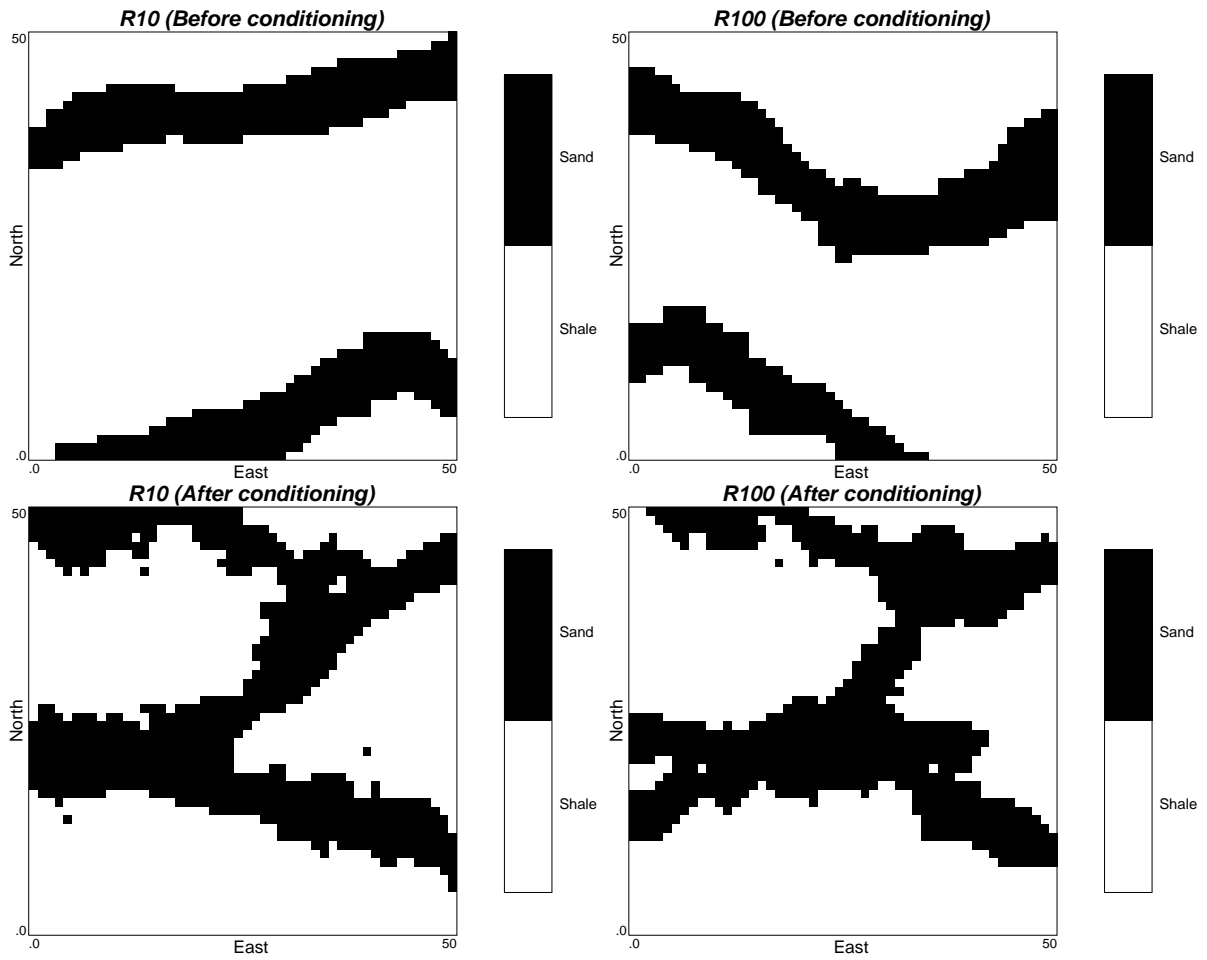


Figure 6: Randomly selected two realizations before and after conditioning

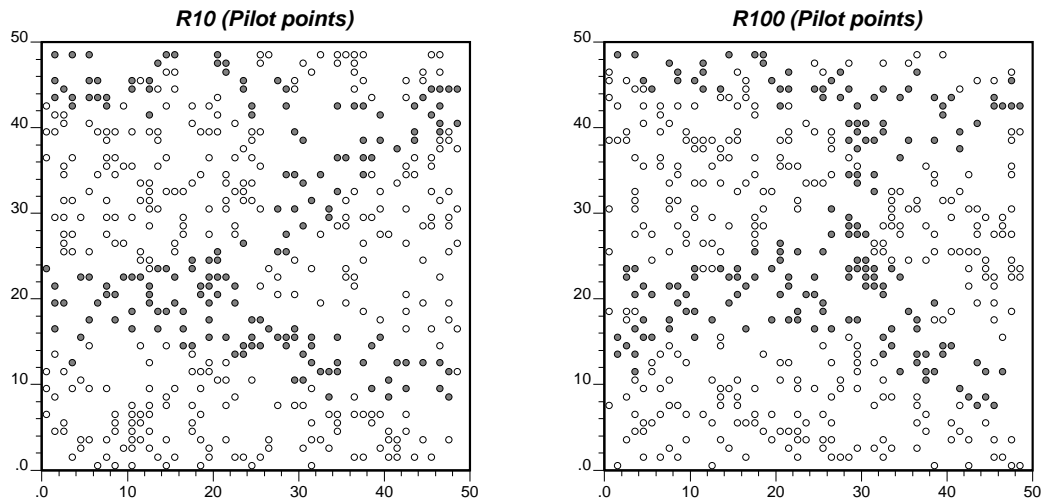


Figure 7: Spatial distribution of pilot points and their estimated values for two realizations (white color means the shale and black color corresponds to sand )

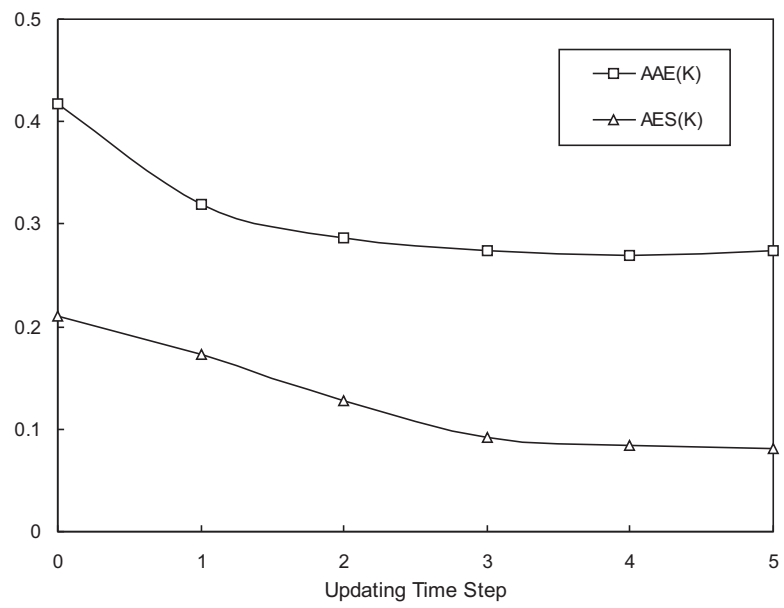


Figure 8: The evolution of conductivity AAE and AES with the updating time steps

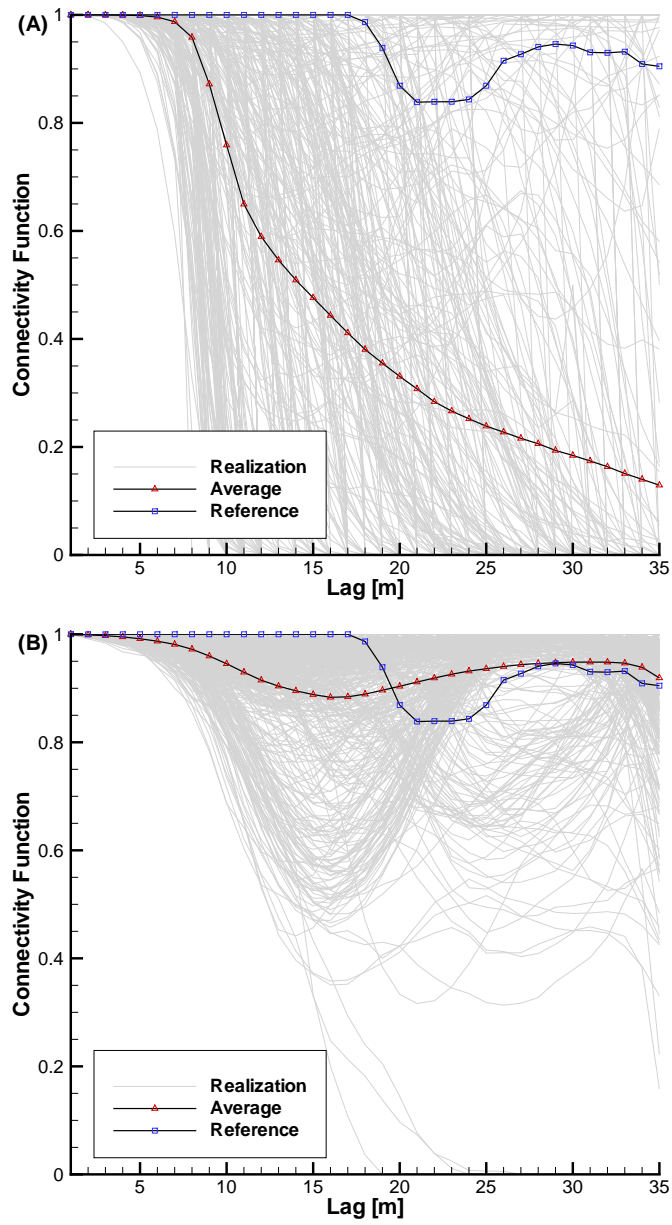


Figure 9: The ensemble connectivity of conductivity before (A) and after (B) data conditioning

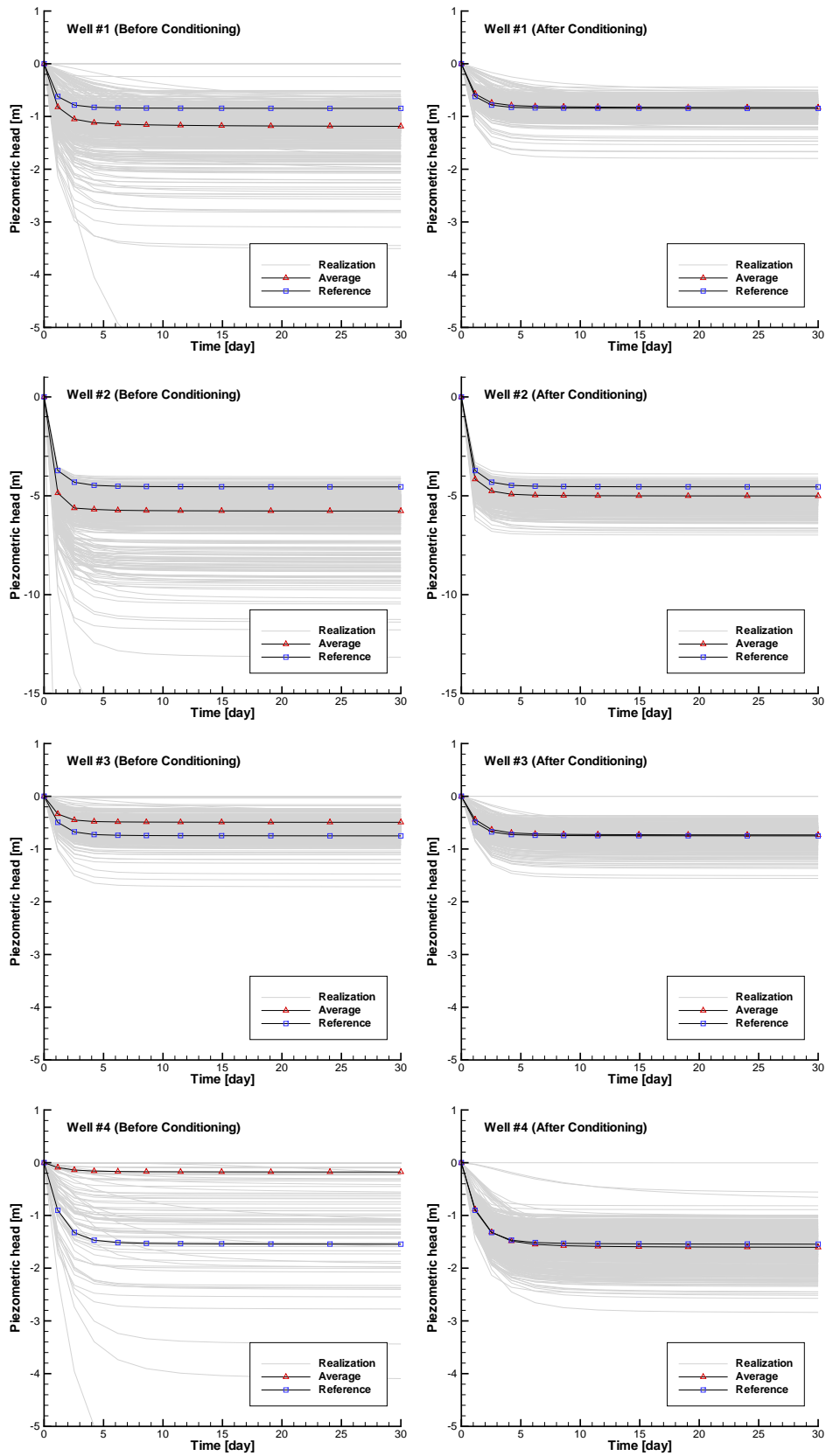


Figure 10: The simulate head before and after data conditioning at 4 wells

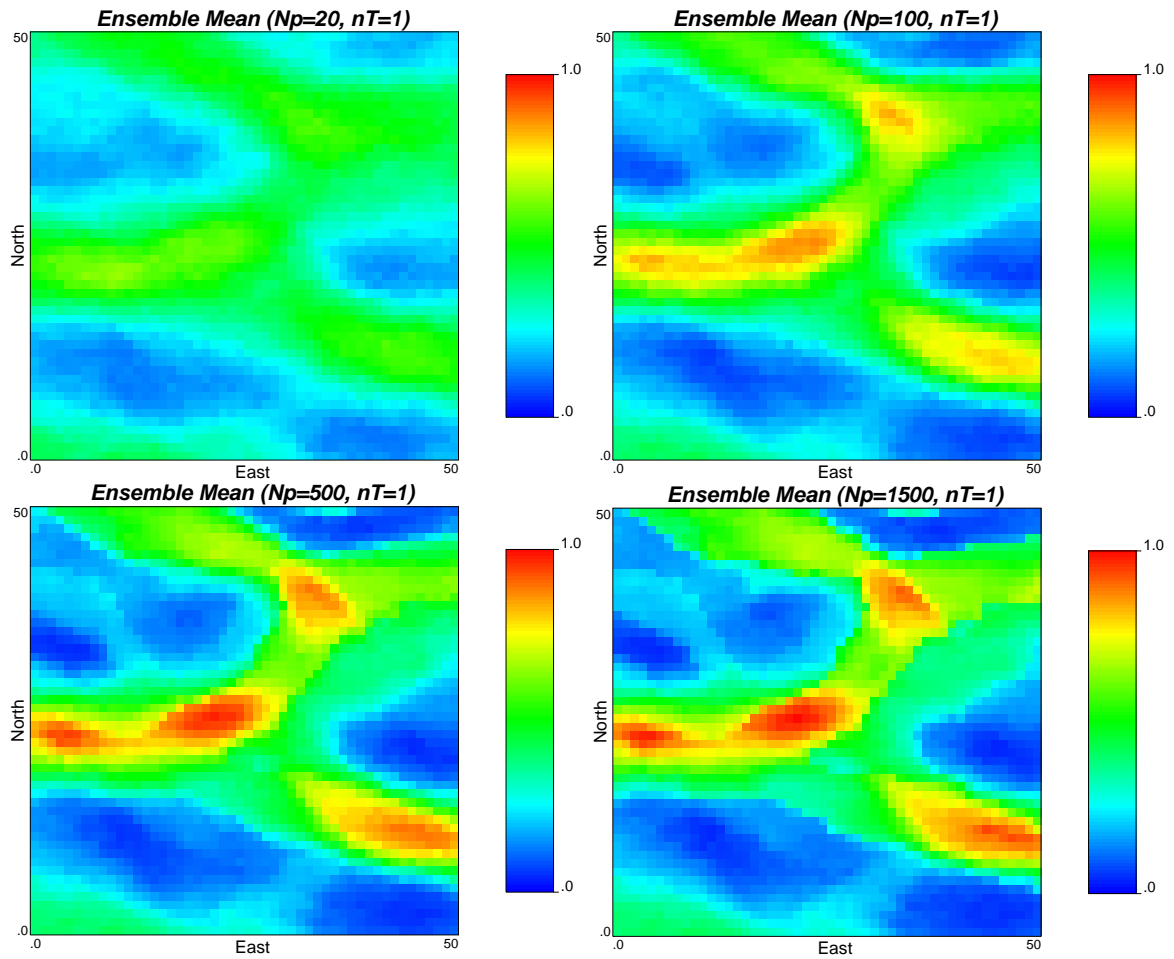


Figure 11: The ensemble mean of conductivity using different number of pilot points



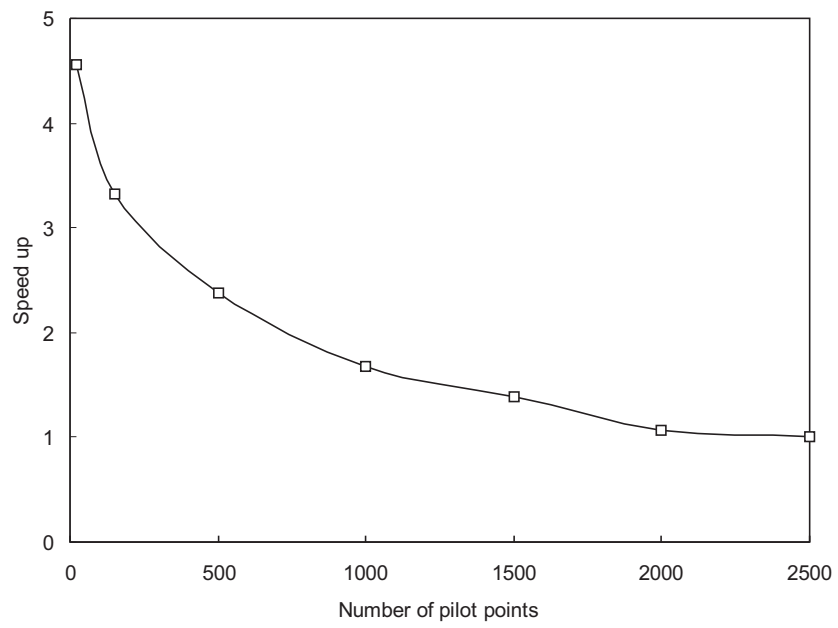


Figure 12: The speed up of the EnPAT algorithm using different number of pilot points