Xu, T.; Gómez-Hernández, JJ. (2015). Probability fields revisited in the context of ensemble Kalman filtering. Journal of Hydrology. 531(1):40-52. doi:10.1016/j.jhydrol.2015.06.062.

Probability Fields Revisited in the Context of Ensemble Kalman Filtering

Teng Xu[a,*], J. Jaime Gómez-Hernández[a]

[a]Research Institute of Water and Environmental Engineering, Universitat Politècnica de València, 46022, Valencia, Spain

Abstract

Hu et al. (2013) proposed an approach to update complex geological facies models generated by multiple-point geostatistical simulation while keeping geological and statistical consistency. Their approach is based on mapping the facies realization onto the spatially uncorrelated uniform random numbers used by the sequential multiple-point simulation to generate the facies realization itself. The ensemble Kalman filter was then used to update the uniform random number realizations, which were then used to generate a new facies realization by multiple-point simulation. This approach has not a good performance that we attribute to the fact that, being the probabilities random and spatially uncorrelated, their correlation with the state variable (piezometric heads) is very weak, and the Kalman gain is always small. The approach is reminiscent of the probability field simulation, which also maps the conductivity realizations onto a field of uniform random numbers; although the mapping now is done using the local conditional distribution functions built based on a prior statistical model and the conditioning data. Contrary to Hu et al. (2013) approach, this field of uniform random numbers, termed a probability field, displays spatial patterns related to the conductivity spatial patterns, and, therefore, the correlation between probabilities and state variable is as strong as the correlation between conductivities and state variable could

*Corresponding author. Tel: +34 963879615 Fax: +34 963879492

Email addresses: tenxu@posgrado.upv.es (Teng Xu), jaime@dihma.upv.es (J. Jaime Gómez-Hernández)

be. Similarly to Hu et al. (2013), we propose to use the ensemble Kalman filter to update the probability fields, and show that the existence of this correlation between probability values and state variables provides better results.
Keywords: MPS; non-Gaussian; sequential simulation; inverse modeling

1. Introduction

The ensemble Kalman filter (EnKF) Evensen (1994, 2003), is an effective and computationally efficient data assimilation method, which has received much attention in the inverse modeling community, since it can be applied for the inversion of the parameters controlling a non-linear state-transfer function given some state observational data. The EnKF is optimal for the case in which the state-transfer function is linear and parameters and state variables are multiGaussian (Aanonsen et al., 2009); it has proven to work remarkably well for non-linear state-transfer functions; but it has failed when trying to deal with non-Gaussian fields (Simon and Bertino, 2009; Chen et al., 2009; Sun et al., 2009). Recently, several methods have been developed trying to handle non-Gaussianities in EnKF; for example, those combining the EnKF and a Gaussian mixture model (Franssen and Kinzelbach, 2008; Gu and Oliver, 2007), those using iterative EnKF (Franssen and Kinzelbach, 2008; Gu and Oliver, 2007), those combining the EnKF with Gaussian anamorphosis (also referred as normal-score transform) (Bertino et al., 2003; Béal et al., 2010; Zhou et al., 2011; Xu et al., 2013b) and those combining the EnKF with multiple-point geostatistics (Hu et al., 2013).

Considering that the strength of multiple-point geostatistics is dealing with non-Gaussian fields (Guardiano and Srivastava, 1993; Strébelle, 2000), Hu et al. (2013) proposed a method to use the EnKF with non-Gaussian reservoir models by mapping facies onto the uniform random numbers used to generate them. In multiple-point geostatistical simulation (MPS), realizations are generated using the sequential simulation principle (Gómez-Hernández and Journel, 1993), whereby each node of the grid is visited, a local conditional distribution is

constructed, and then a uniform random number is generated that is used to draw a value from the conditional distribution. There is a unique relationship between the (independent) uniform random numbers and the attribute values; therefore, one can envision using the random numbers as the parameters to be updated by the EnKF algorithm, and thus preserving the non-Gaussian features that are built into the calculation of the conditional distributions. The idea is very clever because once you fix all other parameters in the MPS algorithm, that is, training image, size of search neighborhood to look for conditioning data, maximum number of conditioning data to retain, path for the sequence in which the nodes are visited, etc., you can modify locally or globally the field of uniform random numbers to generate a new reservoir model.

The initial objective of the work by Hu et al. (2013) was to assimilate production data onto binary facies models; the mapping of the uniform probability realization onto a facies realization (a realization consisting of only two numbers) has the additional interest of finding a mapping of a discrete field onto a continuous one, since the latter will be amenable of treating by the EnKF. The method proposed by Hu et al. (2013) simply applies the standard EnKF to the uniform random numbers, instead of onto the facies values.

We have tested the method by Hu et al. (2013) in the context of assimilating piezometric heads in an aquifer and we have found that the method does not perform as well as expected, at least for the case analyzed hereafter. We think that this underperformance is due to the very weak cross-correlation that there is between the uniform numbers and the state variables. Recall that the EnKF proceeds in two steps: forecast and analysis. The forecast step presents no problem, it is based in the solution of the numerical model appropriate to the process being studied. The analysis step is the one in which the approach by Hu et al. (2013) fails. In the analysis step, discrepancy between predicted and observed states at observation locations is used to update the parameters driving the state-equation. This update is proportional to the said discrepancy, but also to what is called the Kalman gain,

which is a function of the auto- and cross-covariances of parameters and state variables. When the parameters being updated are uniform random numbers that are uncorrelated in space, the auto-covariance of the parameters and the cross-covariance are very weak, resulting in a very small Kalman gain. The net effect is that during the analysis step the update of the uniform random field is small and limited to a very narrow area around observation locations.

Mapping parameters onto probabilities reminded us of the probability field approach Froidevaux (1993) to generate conditional realizations of a given parameter using unconditional (but correlated) realizations of a uniform random field. In this case, the mapping uses the local conditional distribution functions of the parameter. Given a set of conditional data, and a set of structural parameters, one can obtain the local conditional distribution functions by simple kriging, indicator kriging, numerically from a training image, or from an ensemble of realizations generated otherwise. Once the local conditional distribution functions are defined at each point within the domain, there is a unique mapping from a probabilty field onto a parameter field. The probability field does not have to be conditional to the parameter values since the conditioning will happen when reading back the (Heaviside) cumulative distribution function at conditioning locations; yet, it needs to be correlated, to preserve the correlation structure of the parameters. The interest of the probability field approach was the generation of conditional realizations (of the parameter) from unconditional realizations (of probabilties), which were, at the time, much cheaper to generate than conditional ones. The method never had a wide acceptance for the difficulty of establishing, a priori, which the correlation structure of the probabilities should be. Yet, there are some interesting applications of probability fields for inverse modeling (Capilla et al., 1999; Capilla and Llopis-Albert, 2009).

We have decided to revisit the concept of probability fields in the context of data assimilation by the EnKF. The spatial correlation of the probability fields will be determined a

posteriori, from the ensemble of parameter realizations, thus avoiding the main problem of the original idea. We have applied the EnKF on the probability fields and found that the method gives good results.

The paper continues by presenting an extension to the algorithm by Hu et al. (2013) for the simulation of continuous variables, together with the implementation of the EnKF using probability fields. Next, both algorithms are tested in a synthetic channelized aquifer with a bimodal histogram of conductivities.

2. Methodology

In this section, we will describe the two algorithms. The algorithm based on Hu et al. (2013) work will be referred to as the Uncorrelated Probability Field (UPF) method, and the one based on Froidevaux (1993) approach will be referred as the Correlated Probability Field (CPF) method.

In both methods, the parameter which is updated by the EnKF is probability (which should follow a uniform distribution between 0 and 1); however, given that the EnKF is optimal when the parameters follow a Gaussian distribution, we will convert the uniform probabilities into Gaussian deviates and apply the EnKF to the latter. The conversion simply amounts to replacing each uniform deviate onto the Gaussian deviate read from the standard $\mathcal{N}(0,1)$ Gaussian cumulative distribution function.

2.1. Uncorrelated probability field method

Hu et al. (2013) approach only updated the facies distribution and then assumed homogeneous parameters within each facies. We would like to have heterogeneous conductivities on top of the facies heterogeneity. For this reason, the generation of each conductivity realization is done in two steps, in the first step a facies realization is generated using MPS (we limit this analysis to two facies), and, in the second step, each facies is independently populated

with conductivity values using sequential Gaussian simulation (SGS) Gómez-Hernández and Cassiraga (1994); Gómez-Hernández and Journel (1993). This implies that the mapping between conductivities and probabilities requires not just one random field of uncorrelated uniform numbers, but three, one to generate the facies and two to generate the conductivities that will be associated with each facies. Conceptually the approach is the same as the original one: conductivities are mapped onto uncorrelated probabilities, and these probabilities are the parameters updated by the EnKF.

Figure 1 shows the flowchart for this method, which can be described as follows:

1. Generate an ensemble of uncorrelated standard Gaussian fields $\mathcal{N}(0,1)$.
2. Transform the uncorrelated Gaussian deviates into uniform probability fields.
3. Use these uniform fields with MPS and SGS to generate an ensemble of conductivity fields.
4. Feed the conductivity fields into the groundwater flow model, with appropriate initial and boundary conditions, and given sinks and sources; as a result, we obtain an ensemble of piezometric heads.
5. Compute the auto- and cross-covariances of the Gaussian deviates obtained in step 1 and the piezometric heads of the previous step.
6. Sample the piezometric heads at the observation points.
7. Update the Gaussian deviates by the EnKF and return to step 1 (we use the standard implementation of the EnKF as proposed initially by Evensen (2003) and described, for instance by Chen and Zhang (2006); Xu et al. (2013a)).

2.2. Correlated probability field method

For the CPF method we need to establish first which are the local conditional probability distributions given the conditioning data. When the probability field method was developed, these probability distributions were obtained by kriging: simple kriging when the random

function model was multiGaussian, indicator kriging when the model was non-parametric and based on the indicator covariances. With the advent of multiple-point geostatistics, model statistics are not specified analytically but inferred from a training image; the training image can also be used to define the local conditional probability distributions. An alternative, used in this work, is to infer the conditional distributions from an ensemble of realizations that has been generated by whichever stochastic simulation approach.

Before the CPF method starts, we have to generate an initial ensemble of conductivity realizations from which to compute the local conditional distribution function at each grid node. (When there are no conditioning data, the local conditional distribution coincides with the marginal one everywhere.)

Considering that the map of local distribution functions has been already determined by one way or another, the flow chart of the CPF is shown in Figure 2, which can be described as follows:

1. Transform the ensemble of conductivity fields into an ensemble of (correlated) probability fields by replacing each conductivity value with the probability associated to its local conditional distribution.
2. Transform the probabilities into Gaussian deviates using an inverse standard Gaussian cumulative distribution function $\mathcal{N}(0,1)$.
3. Feed the conductivity fields into the groundwater flow model, with appropriate initial and boundary conditions, and given sinks and sources; as a result, we obtain an ensemble of piezometric heads.
4. Compute the auto- and cross-covariances of the Gaussian deviates obtained in step 2 and the piezometric heads of the previous step.
5. Sample the piezometric heads at the observation points.
6. Update the Gaussian deviates by the EnKF.

7. Transform the Gaussian deviates into updated probabilities using the standard Gaussian cumulative distribution function.
8. Transform back the updated probability fields into updated conductivities through the inverse local conditional distribution and return to step 1.

3. Synthetic Example

The performance of the two methods will be evaluated on a synthetic confined aquifer of 50 m by 50 m by 5 m, discretized into 50 by 50 by 1 cells. The aquifer is composed of 35% high conductivity sand and 65% low conductivity shale. The spatial heterogeneity of the sand/shale distribution is characterized by the training image shown in Figure 3 (replicated after (Strebelle, 2002)). Hydraulic log-conductivity within the sand follows a Gaussian distribution with mean of 2.3 ln(m/d) and a standard deviation of 1 ln(m/d), while in shale has a mean of -3.5 ln(m/d) and a standard deviation of 0.6 ln(m/d). The conductivity in sand has an anisotropic spatial correlation characterized by an exponential variogram with ranges of 48 m in the horizontal direction and 24 m in the vertical direction; while conductivity in shale displays an isotropic correlation characterized by an exponential variogram with range of 24 m. These parameters are summarized in Table 1.

The log-conductivities in the synthetic aquifer are built in two steps. In the first step, using the code SNESIM Strebelle (2002), a binary sand/shale realization is generated coherent with the training image in Figure 3. (Note that the training image extends over a much larger area than the aquifer.) In the second step, using the code GCOSIM3D Gómez-Hernández and Journel (1993), each facies is populated independently with log-conductivity values generated by sequential Gaussian simulation. The resulting field and its histogram are shown in Figure 4. The synthetic aquifer displays the channelized structure of the training image and has a bimodal distribution with global mean of -1.1 ln(m/d) and global standard deviation of 2.8 ln(m/d).

A transient groundwater flow problem is solved in the synthetic aquifer using MODFLOW (McDonald and Harbaugh, 1984; Harbaugh et al., 2000). The top and bottom boundaries are impermeable, the left boundary has a prescribed head equal to 8 m, and the right boundary has prescribed pumping at the segments coinciding with the sand channels with a total pumping of 190 m^3/d distributed as shown in Figure 4. The initial head is uniform and equal to 8 m throughout. The specific storage is homogeneous en equal to 0.1 m^{-1}. The total simulation time is 500 d, discretized into 100 time steps following a geometric sequence with ratio 1.05. (The first time step is 0.19 days.)

After solving groundwater flow, the piezometric heads at the 25 points shown in Figure 5 are recorded and saved for all time steps. The data from the first 60 time steps (67 days) will be used for assimilation by the EnKF. No facies data, or log-conductivity data are used.

Both the UPF and the CPF begin with an ensemble of realizations that will be progressively updated by ensemble Kalman filtering after observations are taken at each time step. In this work, we generate 800 realizations, using the same two-step approach as for creating the synthetic aquifer, that is, we use the same training image for the facies realizations by MPS and the same parameters of Table 1 to fill in the facies with log-conductivities. These realizations are unconditional since no data on facies or log-conductivity are available.

For the purpose of applying the UPF method, we have recorded the uniform random numbers used for the generation of the fields; then, these uniform random numbers are transformed into Gaussian deviates using the inverse of the standard Gaussian distribution. The UPF starts from this ensemble of realizations of Gaussian deviates (see Figure 1).

For the purpose of applying the CPF method, we have computed, at each node, the local conditional distribution function, which, for this case, since the realizations are unconditional, coincides with the one derived from the global histogram of the reference (see Figure 4b).

Next, we evaluate the ability of both methods to reproduce the patterns observed in the

synthetic aquifer after assimilating the piezometric heads for the first 60 time steps. We will also evaluate the ability of both methods to reproduce the observed piezometric heads.

4. Reproducing log-conductivity

Figure 6 shows the ensemble mean and the ensemble variance of the initial ensemble of log-conductivity realizations used in both approaches, together with the global histogram. The 800 realizations were generated unconditional and, consequently, their ensemble mean and ensemble variance are flat and equal to their marginal values.

Figures 7 shows the ensemble mean, ensemble variance and global histograms computed for both methods after assimilating piezometric head data for 10 time steps. Figure 8 shows similar results after 60 time steps. From these two figures it is evident that the CPF outperforms the UPF. The ensemble mean of the log-conductivities obtained by the CPF already delineates the channels observed in the synthetic aquifer, and the ensemble variance highlights that there is some residual uncertainty at the channel edges. Whereas the ensemble mean and variance of the log-conductivities obtained by the UPF are a little bit more informative than the mean and variance for the initial ensemble, yet, they are quite far from the results achieved by the CPF.

Notice that the bimodal histogram of the log-conductivities is respected by the ensemble of updated realizations in both methods. In the case of the UPF this happens by construction, since, independently of the updated probabilities resulting from the application of the EnKF, the associated log-conductivities are fully consistent with the prior statistical model since they are obtained by MPS followed by SGS with fixed training image and statistical parameters. In the case of the CPF it depends on the histogram of the probabilities, when this histogram remains uniform; then, sampling back the local conditional distributions will result in a global histogram respecting the prior one.

To try to explain why the dramatic difference in performance we will show the evolution

of realizations number 400 and 800 at different times during the assimilation process. Figures 9 and 10 show these two realizations at the beginning (same realization for both approaches) and after time steps 9, 10, 59 and 60. The objective of this display is to show the changes that can occur in a single update step (between time steps 9 and 10 or between time steps 59 and 60) and the overall evolution. In the UPF, the changes between consecutive time steps can be quite drastic, given the nature of the sequential simulation algorithms, in which each node is simulated based on the previously simulated nodes; however, since each realization is obtained by MPS and SGS, the delineation of the facies (as in the training image) is quite clear at all steps. In the CPF, the changes between consecutive time steps are quite smooth, the EnKF updates smoothly the probabilities, and when reading back the updated probabilities through the local distribution functions the changes are smooth, too; however, the controls that generated the facies in the initial realizations disappear, and the delineation of the facies becomes fuzzier than in the UPF, yet the look of the final updated realizations obtained with the CPF is closer to the reference than with the UPF.

It is interesting to analyze how the update in the log-conductivity field relates to the update of the underlying probability fields. Figure 11 shows the update of the underlying Gaussian deviates at time steps 10 and 60 as computed by the EnKF in realization 400 for both methods. In the case of the UPF, there are three Gaussian deviate fields, we are showing only the field that is used to update the facies distribution. Similar results are shown for realization 800 in Figure 12. The update is the result of the product of the Kalman gain by the discrepancy between predicted and observed piezometric heads. In the UPF, the updates of the Gaussian deviates are very local and short correlated, as a result of the very weak correlation between the uncorrelated probability fields and the piezometric heads; however, this local, random looking update of the probabilities produces quite important changes in the facies. Figure 13 shows the facies change in realizations 400 and 800 at time steps 10 and 60. Apparently, similar random updates of the Gaussian deviates induce quite

different updates in the facies realization for realizations 400 and 800. This "discrepancy" between the Gaussian update and the facies update is due to the nature of the MPS and SGS sequential simulation algorithms, which can be regarded as chaotic. On the contrary, in the CPF, the updates in the Gaussian deviates are smoother than in the previous method and their magnitude reduce significantly as time proceeds. This reduction in the magnitude of the updates is the result of the piezometric head assimilation which in turn reduces the ensemble variance.

To illustrate what we mean by chaotic behavior in the sequential simulation algorithms consider two probability fields that are identical except for one pixel. Each probability field will have a conductivity realization associated. If this pixel is at the beginning of the random path used to generate the realizations, the change in hydraulic conductivity at that location will induce changes in the nearby locations, since all conditional probabilities will be influenced by this initial change. However, if the pixel at which the probability fields differ is the last in the simulation path, only the conductivity at that location will be different between the two conductivity fields. This is shown in Figure 14

5. Reproducing piezometric head

Figure 15 shows the piezometric head evolution at piezometers #1 and #2 of Figure 5 computed on the initial set of realizations. As expected, and given that the initial realizations are unconditional, their response to the groundwater flow conditions in the synthetic aquifer is quite variable among the realizations.

Figure 16 shows the piezometric head evolution at the same piezometers for the two methods after 60 assimilation time steps. The UPF shows a minor improvement with respect to the initial realizations, whereas the CPF is able to generate log-conductivity realizations capable of matching almost perfectly the observed piezometric heads.

Again, the behavior of the UPF must be attributed to the weak correlation between the

Gaussian deviates and the piezometric heads in the UPF, plus the chaotic behavior of the sequential simulation algorithms: a small change in a single probability value could induce a very large change in the final log-conductivity map, particularly if this change happens in a node that is generated early in the path that visits all nodes being simulated.

6. Discussion

The idea of mapping the conductivities onto probabilities was not new, but choosing as probabilities the ones used in the sequential simulation algorithm to draw from the local distribution functions conditioned to all previously simulated values was certainly a new idea. By modifying these probability fields and using them in a multiple-point geostatistical simulation, we can assure that the final realizations will always be coherent with the training image chosen. Therefore, it seemed a good idea to try to update the probability fields, instead of the conductivities directly in the context of the EnKF, in order to apply the EnKF to the generation of clearly non-Gaussian realizations. However, as it has been shown, the method has definite flaws linked precisely to the mapping procedure: on one hand, the probabilities are spatially uncorrelated and display a weak correlation with the state variables, on the other hand, the transformation of the perturbation of the probabilities onto perturbations of conductivities is chaotic, with small probability perturbations possibly inducing very large and widespread perturbations in conductivity.

Revisiting the probability field approach, which is based also in the mapping of conductivities onto probabilities, and formulating the EnKF method in terms of these probabilities, proves to be a powerful approach to generate conductivity realizations which display features difficult to model with multiGaussian-based approaches. This is an approach that should be reconsidered for data assimilation in hydrogeology and petroleum engineering.

Acknowledgements The first author acknowledges the financial support from the China

Scholarship Council (CSC). Part of this work was done while the second author was on sabbatical with the Kansas Geological Survey, Kansas University, Lawrence, KS, USA, which was funded by the Spanish Ministry of Education, Culture and Sports through grant PRX14/00501. Financial support to carry out this work was also received from the Spanish Ministry of Economy and Competitiveness through project CGL2011-23295.

Aanonsen, S., Nævdal, G., Oliver, D., Reynolds, A., Vallès, B., 2009. The ensemble kalman filter in reservoir engineering–a review. SPE Journal 14, 393–412.

Béal, D., Brasseur, P., Brankart, J., Ourmières, Y., Verron, J., et al., 2010. Characterization of mixing errors in a coupled physical biogeochemical model of the north atlantic: implications for nonlinear estimation using gaussian anamorphosis. Ocean Science .

Bertino, L., Hollard, A., Evensen, G., Wackernagel, H., 2003. An ensemble kalman filter for non-gaussian variables, in: EGS-AGU-EUG Joint Assembly, p. 5171.

Capilla, J.E., Llopis-Albert, C., 2009. Gradual conditioning of non-gaussian transmissivity fields to flow and mass transport data: 1. theory. Journal of Hydrology 371, 66–74.

Capilla, J.E., Rodrigo, J., Gómez-Hernández, J.J., 1999. Simulation of non-Gaussian transmissivity fields honoring piezometric data and integrating soft and secondary information. Mathematical Geology 31, 907–927.

Chen, Y., Oliver, D., Zhang, D., 2009. Data assimilation for nonlinear problems by ensemble kalman filter with reparameterization. Journal of Petroleum Science and Engineering 66, 1–14.

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

Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. J. Geophys. Res 99, 143–10.

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

Franssen, H., Kinzelbach, W., 2008. Real-time groundwater flow modeling with the ensemble kalman filter: Joint estimation of states and parameters and the filter inbreeding problem. Water Resources Research 44, W09408.

Froidevaux, R., 1993. Probability field simulation, in: Geostatistics Troia 92. Springer, pp. 73–83.

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

Gómez-Hernández, J.J., Journel, A.G., 1993. Joint sequential simulation of Multi-Gaussian fields, in: Soares, A. (Ed.), Geostatistics Tróia '92, Kluwer Academic Publishers, Dordrecht. pp. 85–94.

Gu, Y., Oliver, D., 2007. An iterative ensemble kalman filter for multiphase fluid flow data assimilation. SPE Journal 12, 438–446.

Guardiano, F.B., Srivastava, R.M., 1993. Multivariate geostatistics: Beyond bivariate models, in: Soares, A. (Ed.), Geostatistics Tróia '92, volume 1, Kluwer. pp. 133–144.

Harbaugh, A., et al., 2000. MODFLOW-2000, the US Geological Survey modular groundwater model: User guide to modularization concepts and the ground-water flow process. US Geological Survey.

Hu, L., Zhao, Y., Liu, Y., Scheepens, C., Bouchard, A., 2013. Updating multipoint simulations using the ensemble kalman filter. Computers & Geosciences 51, 7–15.

McDonald, M., Harbaugh, A., 1984. A modular three-dimensional finite-difference ground-water flow model. Scientific Publications Company.

Simon, E., Bertino, L., 2009. Application of the gaussian anamorphosis to assimilation in a 3-d coupled physical-ecosystem model of the north atlantic with the enkf: a twin experiment. Ocean Sci 5, 495–510.

Strébelle, S., 2000. Sequential simulation drawing structures from training images. Ph.D. thesis. Stanford University. 187pp.

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

Sun, A., Morris, A., Mohanty, S., 2009. Sequential updating of multimodal hydrogeologic parameter fields using localization and clustering techniques. Water Resources Research 45, W07424.

Xu, T., Gómez-Hernández, J.J., Li, L., Zhou, H., 2013a. Parallelized ensemble kalman filter for hydraulic conductivity characterization. Computers & Geosciences 52, 42–49.

Xu, T., Gómez-Hernández, J.J., Zhou, H., Li, L., 2013b. The power of transient piezometric head data in inverse modeling: An application of the localized normal-score enkf with covariance inflation in a heterogenous bimodal hydraulic conductivity field. Advances in Water Resources 54, 100–118.

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

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

Facies	Proportion	Mean [ln [m/d]]	Std.dev [ln [m/d]]	Variogram type	λ_x [m]	λ_y [m]	sill
Sand	0.35	2.3	1.0	exponential	48	24	1
Shale	0.65	-3.5	0.6	exponential	24	24	0.35

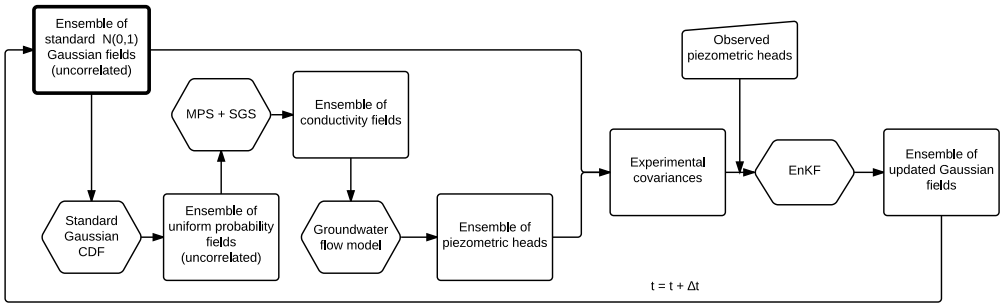

Figure 1: Work flow for the Unconditional Probability Field method. The starting step is highlighted in bold.

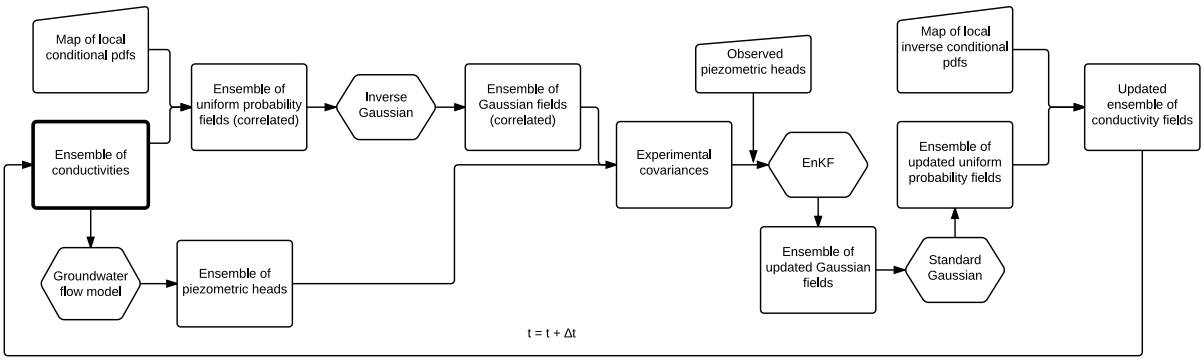

Figure 2: Work flow for the Conditional Probability Field method. The starting step is highlighted in bold.

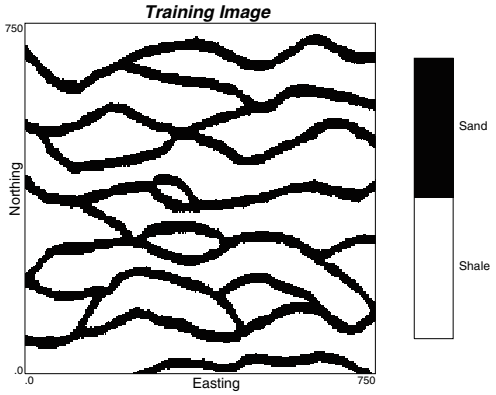

Figure 3: Training image

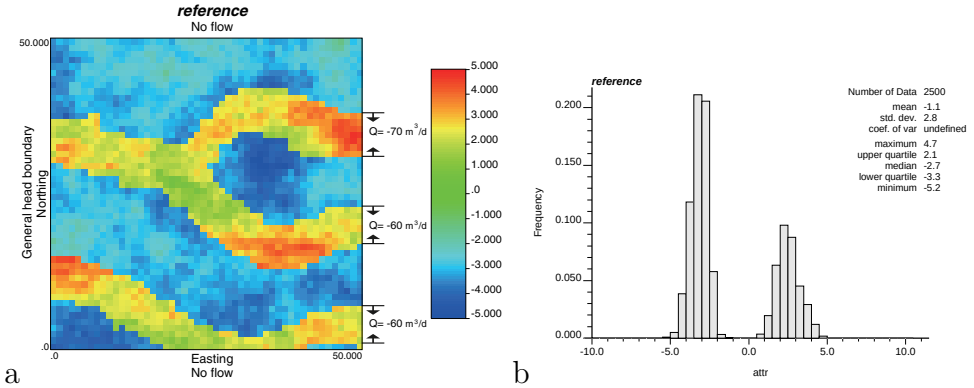

Figure 4: Synthetic aquifer and log-conductivity histogram

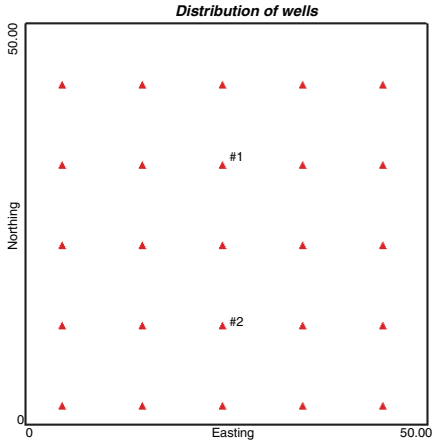

Figure 5: Distribution of observation piezometers

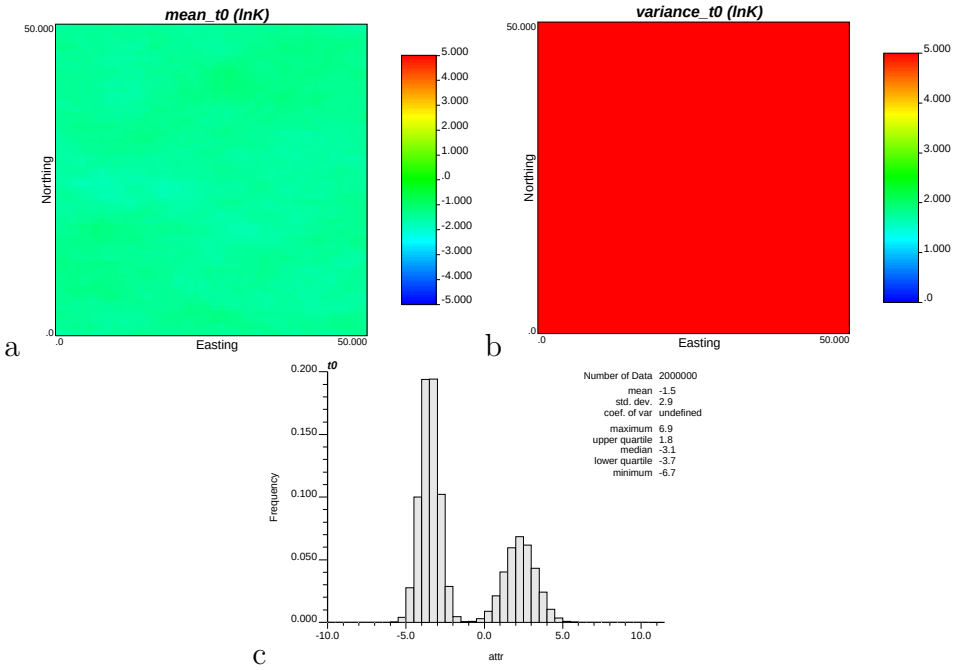

Figure 6: Ensemble mean, ensemble variance and ensemble histogram of the initial ensemble of log-conductivity realizations.

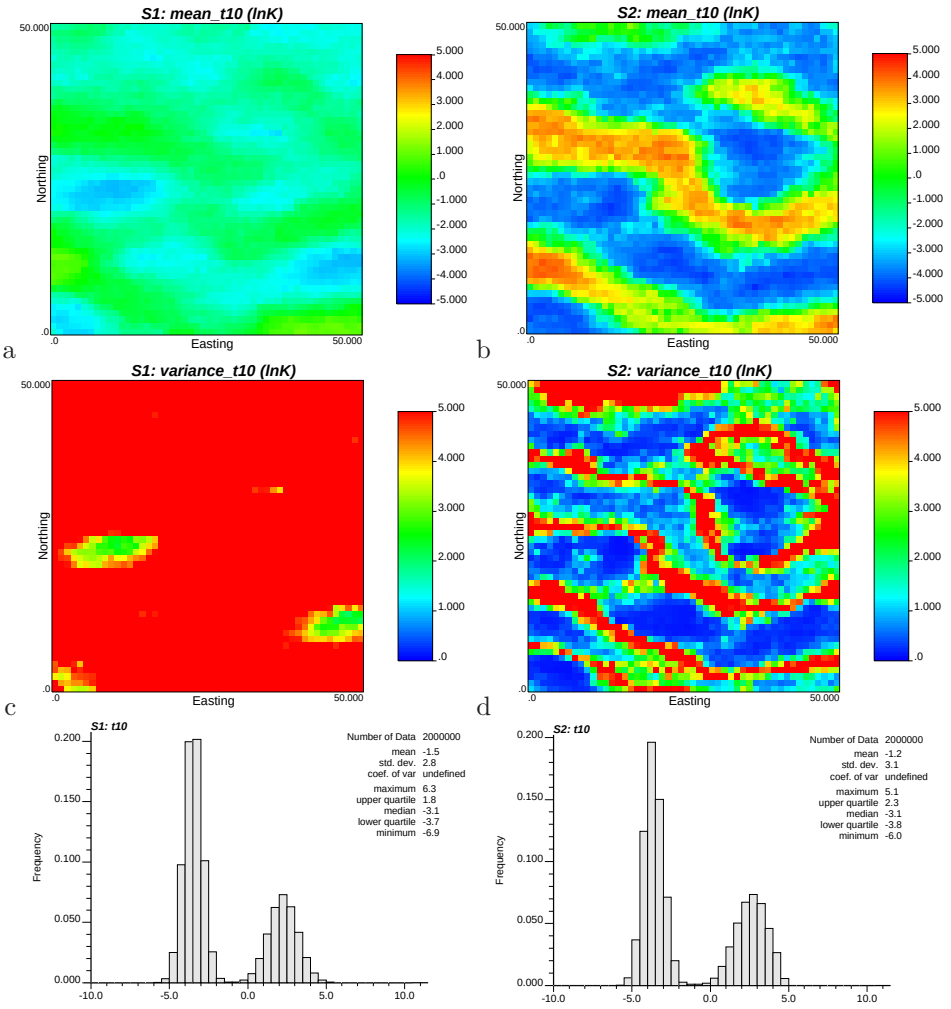

Figure 7: Ensemble mean, ensemble variance and ensemble histogram of the log-conductivity realizations obtained after 10 assimilation time steps. Left column, uncorrelated probability fields (S1). Right column, correlated ones (S2).

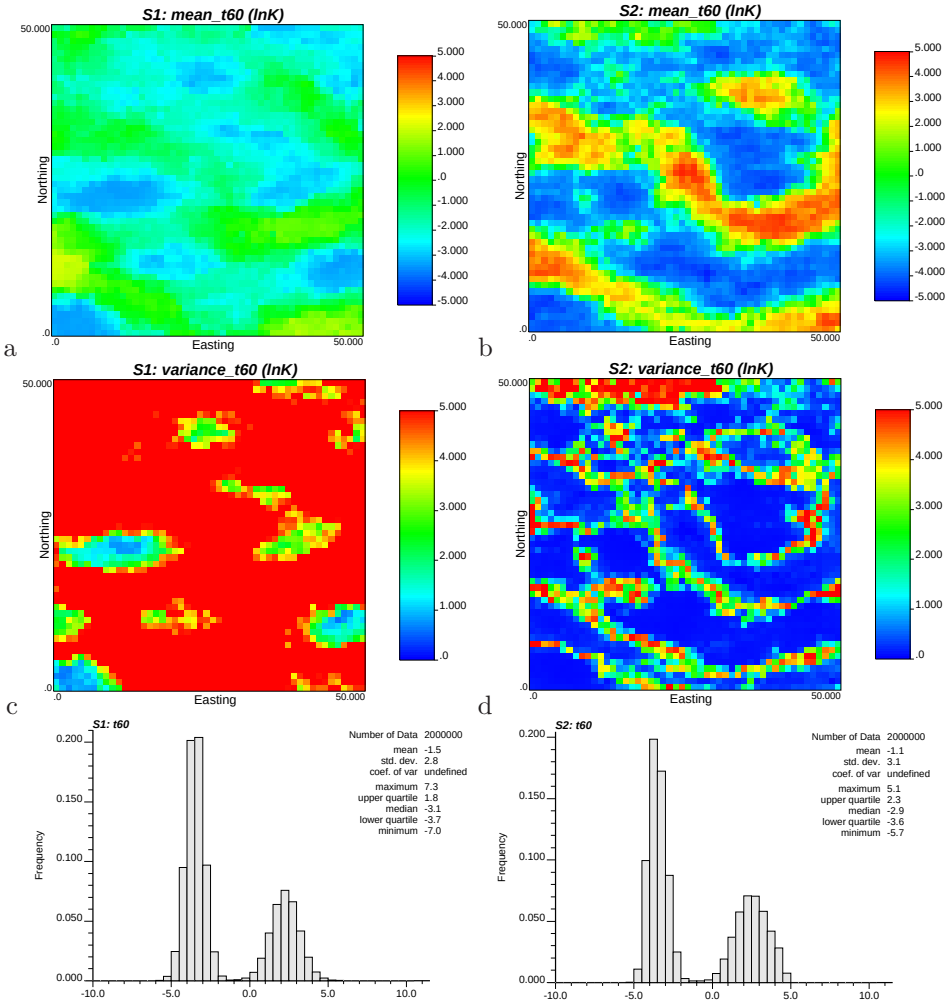

Figure 8: Ensemble mean, ensemble variance and ensemble histogram of the log-conductivity realizations obtained after 60 assimilation time steps. Left column, uncorrelated probability fields (S1). Right column, correlated ones (S2).

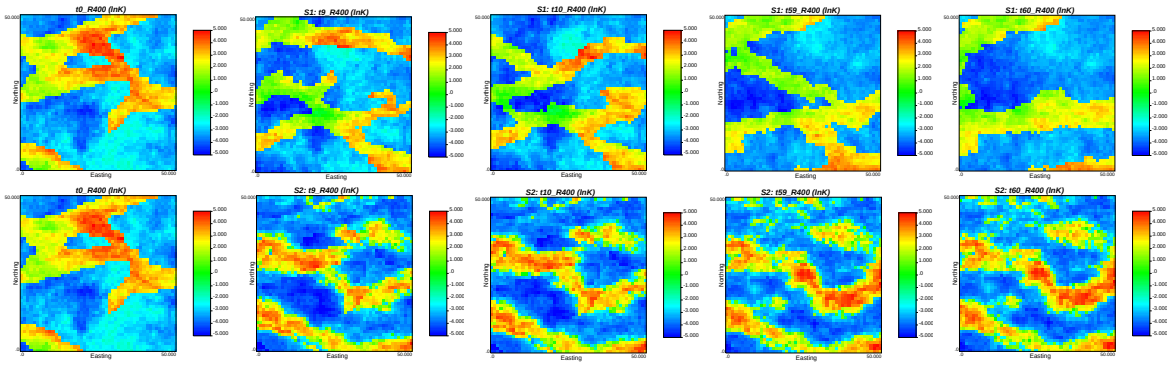

Figure 9: Evolution of the 400th realization in both methods. Top row, uncorrelated probabilities (S1); bottom row, correlated probabilities (S2). Columnwise from left to right, initial realization (same for both approaches), after time step 9, 10, 59 and 60.

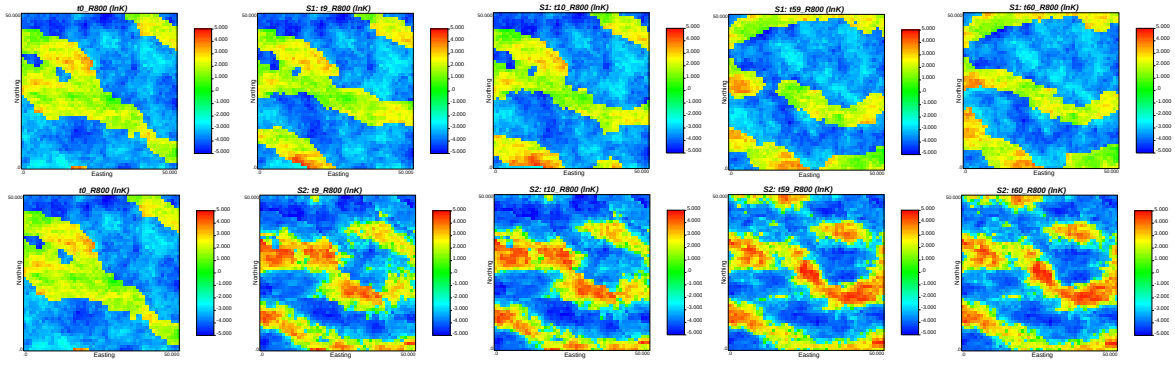

Figure 10: Evolution of the 800th realization in both methods. Top row, uncorrelated probabilities (S1); bottom row, correlated probabilities (S2). Columnwise from left to right, initial realization (same for both approaches), after time step 9, 10, 59 and 60.

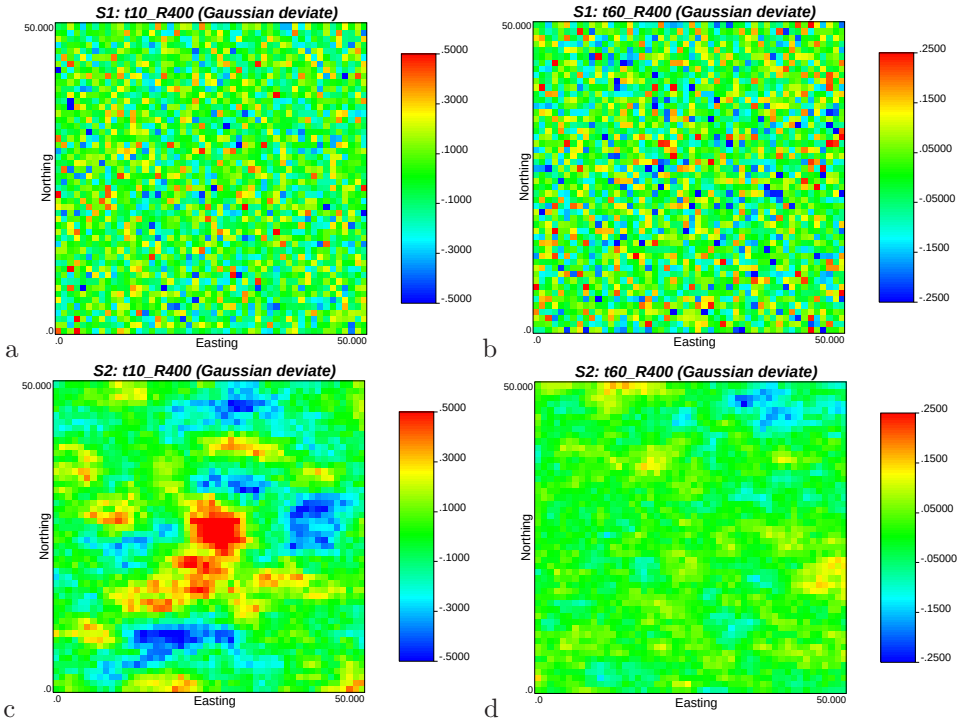

Figure 11: Increment of Gaussian deviates in realization 400. Top row, uncorrelated probabilities (S1); bottom row, correlated probabilities (S2). Left column, update at the 10th step; right column, update at the 60th step.

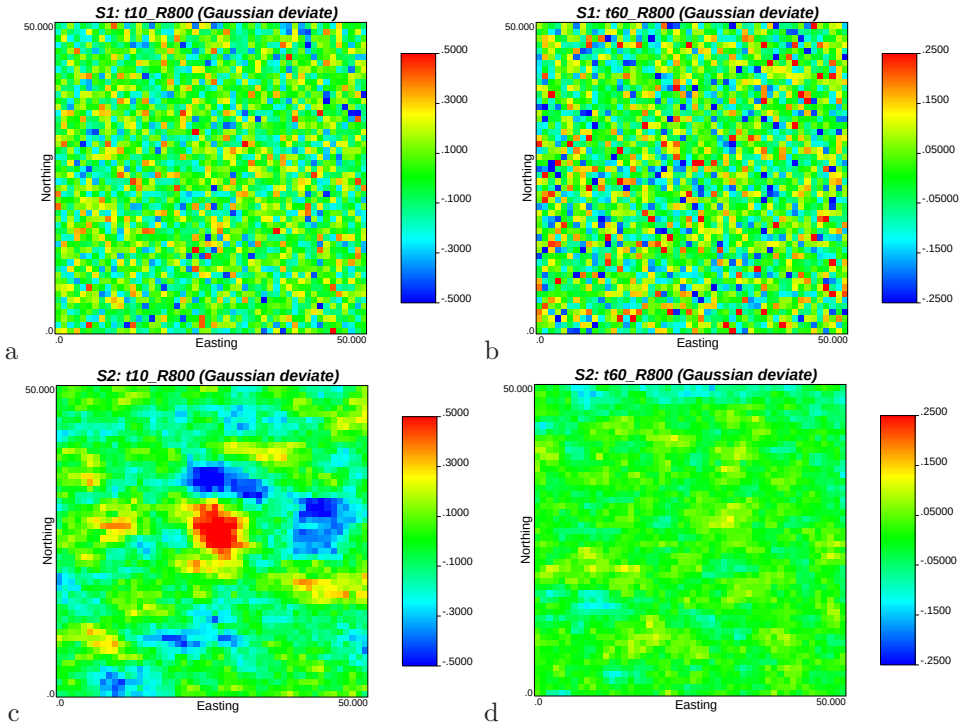

Figure 12: Increment of Gaussian deviates in realization 800. Top row, uncorrelated probabilities (S1); bottom row, correlated probabilities (S2). Left column, update at the 10th step; right column, update at the 60th step.

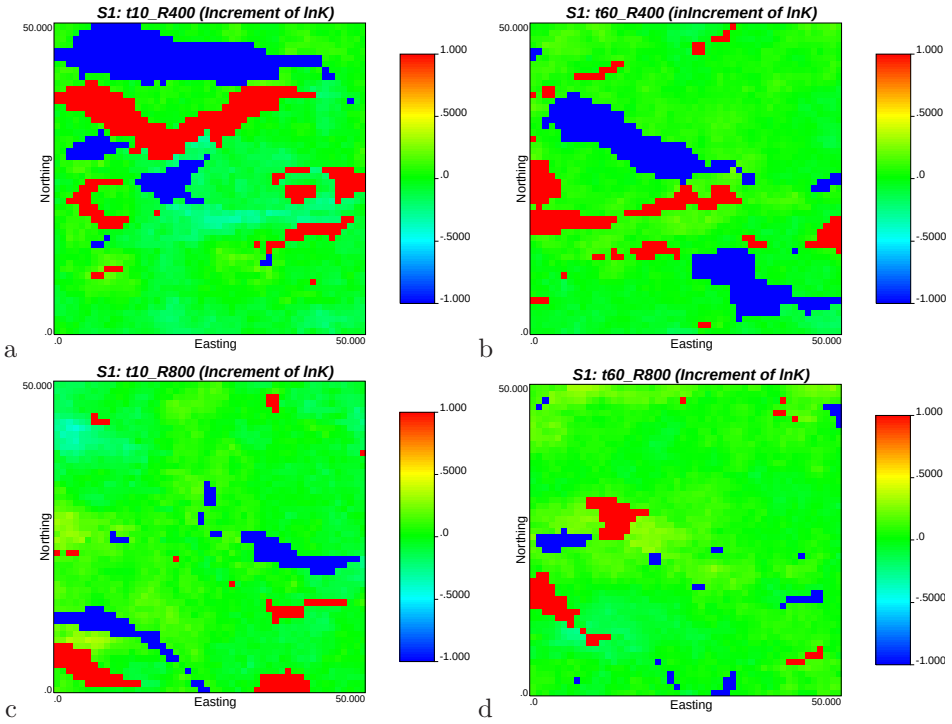

Figure 13: Faces changes induced by updating the probabilities in the uncorrelated probability field method. Blue means change from sand to shale, red means change from shale to sand, and green means no change. Right column for the updates at time step 10, left column for the updates at time step 60, top row for realization 400, bottom row for realization 800.

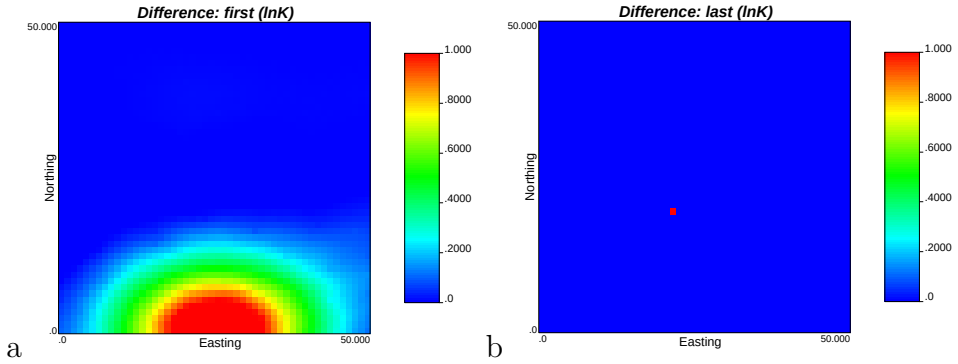

Figure 14: Change in $\ln K$ between two realizations generated by sequential Gaussian simulation using exactly the same parameters except for one of the probabilities used to draw from the local conditional distributions. Left: when the probability changes at the first node of the random path. Right: when the probability changes at the last node of the random path.

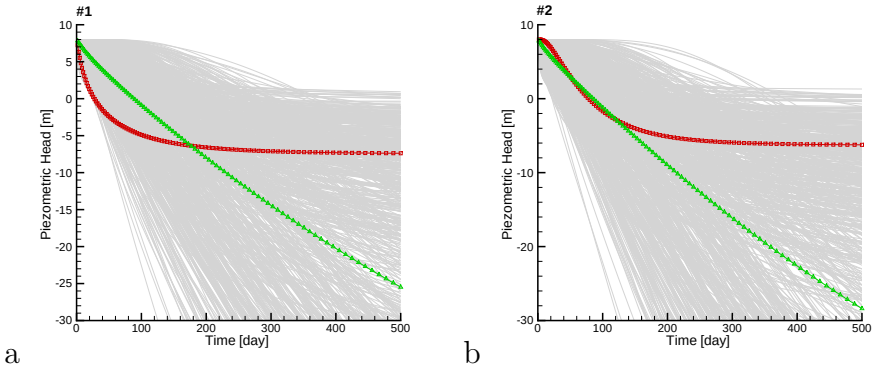

Figure 15: Reproduction of the observed piezometric head at the piezometers #1 and #2 of Figure 5 by the log-conductivity fields obtained after 60 time steps. Left, piezometer #1; right, piezometer #2. The red squares correspond to the head evolution in the synthetic aquifer, the gray lines are the head evolution in the individual realizations, and the green triangles correspond to the ensemble mean.

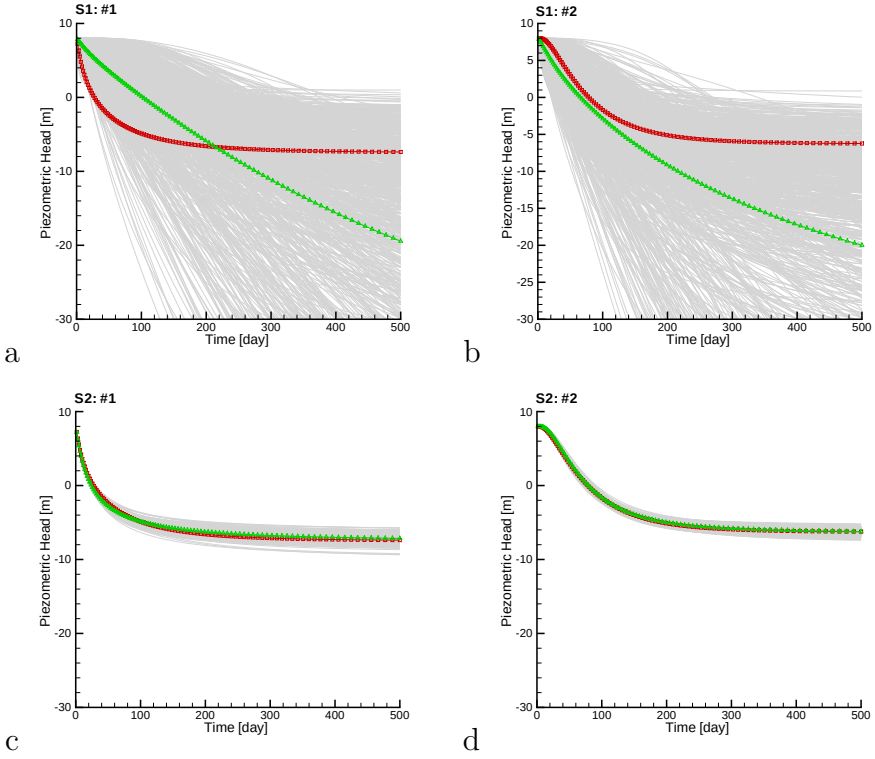

Figure 16: Reproduction of the observed piezometric head at the piezometers #1 and #2 of Figure 5 by the log-conductivity fields obtained after 60 time steps for both methods. Top row, uncorrelated probability field method (S1); bottom row, correlated probability field method (S2). Left, piezometer #1; right, piezometer #2. Meaning of lines same as previous figure.