

Document downloaded from:

<http://hdl.handle.net/10251/46456>

This paper must be cited as:

Zhou ., H.; Li ., L.; HENDRIKUS JOHANNES HENDRICKS-FRANSEN; Gómez-Hernández, JJ. (2012). Pattern Recognition in a Bimodal Aquifer Using the Normal-Score Ensemble Kalman Filter. *Mathematical Geosciences*. 44(2):169-185. doi:10.1007/s11004-011-9372-3.



The final publication is available at

<http://dx.doi.org/10.1007/s11004-011-9372-3>

Copyright Springer Verlag (Germany)

Pattern Recognition in a Bimodal Aquifer using the Normal-Score Ensemble Kalman Filter

Haiyan Zhou^{a,*}, Liangping Li^a, Harrie-Jan Hendricks Franssen^b, J. Jaime Gómez-Hernández^a

^a*Grupo de Hidrogeología, Departamento de Ingeniería Hidráulica y Medio Ambiente, Universitat Politècnica de València, 46022, València, Spain*

^b*Agrosphere, IBG-3, Forschungszentrum Jülich GmbH, 52425, Jülich, Germany*

Abstract

The ensemble Kalman filter (EnKF) is now widely used in diverse disciplines to estimate model parameters and update model states by integrating observed data. The EnKF is known to perform optimally only for multiGaussian distributed states and parameters. A new approach, the normal-score EnKF (NS-EnKF), has been recently proposed to handle complex aquifers with non-Gaussian distributed parameters. In this work, we aim at investigating the capacity of the NS-EnKF to identify patterns in the spatial distribution of the model parameters (hydraulic conductivities) by assimilating dynamic observations in the absence of direct measurements of the parameters themselves. In some situations, hydraulic conductivity measurements (hard data) may not be available, which requires the estimation of conductivities from indirect observations, such as piezometric heads. We show how the NS-EnKF is capable of retrieving the bimodal nature of a synthetic aquifer solely from piezometric head data. By comparison with a more standard implementation of the EnKF, the NS-EnKF gives better results with regard to histogram preservation, uncertainty assessment and transport predictions.

Keywords: Large heterogeneity; parameter identification; non-multiGaussian; uncertainty; groundwater modeling; hard data.

*Corresponding author

Email addresses: haizh@upvnet.upv.es (Haiyan Zhou), liali@upvnet.upv.es (Liangping Li), h.hendricks-franssen@fz-juelich.de (Harrie-Jan Hendricks Franssen), jaime@dihma.upv.es (J. Jaime Gómez-Hernández)

1. Introduction

The inverse problem in hydrogeology involves characterizing model parameters, mainly hydraulic conductivity, by integrating measurements of the state variables such as hydraulic head or concentration data. An inverse method that has attracted much attention in the last years is the ensemble Kalman filter (EnKF). The EnKF was proposed by Evensen (1994) and further clarified by Burgers et al. (1998) as an extension to the Kalman filter for the cases in which the state equation is non-linear. It has gained popularity in many disciplines as an efficient data assimilation algorithm (e.g., Houtekamer and Mitchell, 2001; Bertino et al., 2003), and it has been extended, through the augmentation of the state vector, to the identification of the parameters controlling the state of the system (Naevdal et al., 2005; Chen and Zhang, 2006; Moradkhani et al., 2005; Wen and Chen, 2006; Hendricks Franssen and Kinzelbach, 2008). The popularity of the EnKF can be attributed to the relative limited CPU time needed as compared with other Monte-Carlo type inverse modeling (Hendricks Franssen and Kinzelbach, 2009) and the ease of combining the EnKF with virtually any forward model.

The EnKF is known to provide an optimal solution when the state vector follows a multiGaussian distribution and the state transfer function is linear (e.g., Arulampalam et al., 2002). However, in many practical applications of groundwater modeling, hydraulic conductivity cannot be modeled as multiGaussian distributed. The importance of accounting for non-multiGaussinity and the impact of not accounting for it has been clearly demonstrated (e.g., Journel and Deutsch, 1993; Gómez-Hernández and Wen, 1998; Zinn and Harvey, 2003). Zhou et al. (2011) developed a new approach, the normal-score EnKF (NS-EnKF), to better handle non-Gaussian distributions of hydraulic conductivity and states. Both, states and parameters, are normal-score transformed at each time step so that they follow marginal univariate Gaussian distributions. Then, the EnKF is applied on the normal-score transformed states and parameters. The performance of the NS-EnKF was shown to outperform the standard EnKF in a synthetic test example (Zhou et al., 2011). Similar measures have been taken in other disciplines to cope with non-Gaussinity (e.g., Gu and Oliver, 2006; Schoniger and Nowak, 2011; Simon and Bertino, 2009; Wen and Chen, 2007)

The purpose of this paper is to evaluate the performance of the NS-EnKF in identifying hydraulic conductivity patterns in a bimodal aquifer by integrating dynamic hydraulic head data. The NS-EnKF has proven to work well in the presence of a relatively dense sample of hydraulic conductivities, i.e., direct measurements or hard data (Zhou et al., 2011). However, in practice, we have to confront the situation in which the direct measurements of hydraulic conductivity are scarce or simply unavailable. Therefore, we want to investigate the performance of the NS-EnKF to characterize a non-multiGaussian hydraulic conductivity

distribution without the use of hard data. Transient hydraulic heads serve as the only information source for the characterization of hydraulic conductivity; direct measurements of hydraulic conductivity are not considered in this experiment.

The continues as follows. The NS-EnKF algorithm is briefly introduced in Section 2. The method is tested on a synthetic bimodal aquifer composed of sand and shale, where the model is conditioned on the observed hydraulic heads through the NS-EnKF and the standard EnKF. Results and discussions follow the synthetic example. The paper ends with some conclusions.

2. The normal-score ensemble Kalman filter

The standard EnKF algorithm is described by Evensen (1994), Burgers et al. (1998) and Evensen (2007). The first application of the augmented EnKF in hydrogeology was done by Chen and Zhang (2006). Details about the NS-EnKF algorithms can be found in Zhou et al. (2011). The main steps of the NS-EnKF algorithm are the same as for the standard EnKF, but the big difference is the introduction of additional pre- and post- processing steps carried out on the states and parameters contained in the augmented state vector. It starts with an ensemble of realizations that have been generated following a given (non-multiGaussian) random function. The NS-EnKF method consists of the following four main steps:

1. Forward simulation. For each realization of the ensemble, the state vector at time $t - 1$ is updated to time t using a state transfer equation, in our case, piezometric heads are updated from time $t - 1$ to time t through the solution of the groundwater flow equation, the hydraulic conductivities, which are members of the augmented state vector, remain unchanged through this step.
2. Normal score transformation. At each grid cell, and for each component of the augmented state vector (in our case, hydraulic conductivity and piezometric head) compute, using the ensemble of realizations, the local cumulative distribution function (CDFs). These local CDFs are used to construct normal-score transform functions that are used to transform the augmented state vector into a new vector. All the components of the new vector follow a marginal Gaussian distribution with zero mean and unit variance.
3. Update. The new state vector is updated similarly as in the standard EnKF:

$$\mathbf{x}_t^u = \mathbf{x}_t^f + \mathbf{G}_t(\mathbf{z}_t + \varepsilon - \mathbf{H}\mathbf{x}_t^f) \quad (1)$$

where \mathbf{x}_t^u is the vector with the updated state variables at time t (in Gaussian space) and \mathbf{x}_t^f is the

vector computed from the forward simulation and then normal-score transformed; \mathbf{G}_t is the Kalman gain, derived on the basis of the minimization of the posterior error covariance; \mathbf{z}_t is the normal-score transformed observation at time t ; ε is an observation error characterized by a normal distribution with zero mean and a diagonal covariance (it is assumed that errors at different measurement locations are independent); \mathbf{H} is the observation matrix.

4. Back transform. Each component of the updated state vector \mathbf{x}_t^u is back transformed using the previously constructed local CDFs.

The above steps loop until all the observed piezometric heads (for a certain time period) are assimilated.

3. Synthetic example

A synthetic bimodal aquifer composed of sand and shale occupies an area of $300 \text{ m} \times 240 \text{ m}$, where the hydraulic conductivities are characterized by a nonGaussian distribution spreading over several orders of magnitude. The study domain is discretized into 100 columns by 80 rows (the grid cells have dimensions of $3 \text{ m} \times 3 \text{ m}$). The aquifer is assumed confined with a thickness of 10 m.

The reference facies field is generated by SNESIM (Strebelle, 2002), a multiple-point geostatistical simulation algorithm, using as training image the one in Figure 1. Each facies is then populated with log-conductivity values generated by GCOSIM3D (Gómez-Hernández and Journel, 1993) with parameters shown in Table 1. The reference $\ln K$ field is presented in Figure 2. The histogram of log-conductivity is bimodal, with modes coinciding with the means of the sand and shale distributions, $3.0 \ln(\text{m/d})$ and $-2.0 \ln(\text{m/d})$, respectively, and it has a global mean of $-0.47 \ln(\text{m/d})$ and a global standard deviation of $2.39 \ln(\text{m/d})$.

The groundwater flow equation is solved for the reference field using the transient flow simulator MODFLOW (Harbaugh et al., 2000) with impermeable boundary conditions in the north and south, constant head of 0 m in the west and prescribed flow rate of $270.5 \text{ m}^3/\text{d}$ along the east boundary distributed as shown in Figure 2. Different flow rates are set on the eastern boundaries depending on their water supply capacity, i.e., large flow rates correspond to zones with high conductivities. The initial head over the domain is 0 m. The total simulation period of 500 days is discretized into 100 steps with step sizes that increase following a geometric series with a ratio of 1.05. Specific storage is assumed constant and equal to be 0.003 m^{-1} .

Using the same training image (Fig. 1) and the same multiple-point geostatistical algorithm (SNESIM), 1000 facies realizations are generated. For each realization both facies are populated with $\ln K$ values generated by sequential Gaussian simulations using the parameters as specified in Table 1. The assimilation of piezometric head data is performed by the standard EnKF and the NS-EnKF. Observed piezometric

heads from 111 piezometers in 60 time steps (67.7 days) serve as conditioning data. The locations of the 111 piezometers are shown in Figure 2.

4. Results and discussions

4.1. Characterization of $\ln K$ spatial heterogeneity

Figure 3 shows the evolution of the ensemble mean and variance of log-conductivity as the piezometric head data are assimilated by the NS-EnKF from the initial time step ($t = 0$) to the end of the data assimilation period ($t = 60$). The ensemble mean of the initial (unconditional) realizations is flat and shows no feature even though each initial realization honors the same multiple-point geostatistical model implicit in the training image. The structure of the spatial patterns of $\ln K$ starts to appear during the conditioning to piezometric head data. For instance, at the 5th assimilation step, the channel pattern near the eastern boundary is identified and at the end of the 60th step, the ensemble mean of $\ln K$ resembles closely the reference field, having identified quite precisely the channel locations in the reference. The variance is also reduced over the domain as more piezometric heads are integrated. At the end of time step 60 we can even identify the boundaries of the channels by the largest variance strings. We can argue that assimilation of transient hydraulic heads with the NS-EnKF plays a critical role in recognizing patterns of $\ln K$ and allows a good characterization of the bimodal aquifer.

Figure 4 shows the results obtained using the standard EnKF. The ensemble mean is similar to the one obtained by the NS-EnKF, i.e., the $\ln K$ patterns are recognized as more piezometric heads are assimilated. The main difference resides in the variance field. The variance is initially reduced around the piezometers, which is clearly illustrated by observing the variance field up to the 15th time step, while in the case of NS-EnKF, the influence area of the piezometer is extended and depends on the channels to a large extent. Then, at the end of data assimilation ($t = 60$), the $\ln K$ variance in Figure 4 is widely reduced over the entire domain. On one hand, the variance reduction is what we expect as additional information is incorporated in the model; on the other hand, the over-reduction of the ensemble variance cannot represent the real uncertainty, that is, the uncertainty is underestimated by the standard EnKF. This can be reinforced by Figure 5, which shows the evolution of the average absolute bias ($AAB(x)_t$) and the average ensemble spread ($AESP(x)_t$).

The $AAB(x)_t$ and the $AESP(x)_t$ evaluate accuracy and uncertainty of the estimation, respectively. They

are defined as

$$\begin{aligned}
 AAB(x)_t &= \frac{1}{N_b} \sum_{i=1}^{N_b} \frac{1}{N_e} \sum_{r=1}^{N_e} |x_{t,i,r} - x_{ref,i}| \\
 AESP(x)_t &= \left(\frac{1}{N_b} \sum_{i=1}^{N_b} \sigma_{x_{t,i}}^2 \right)^{1/2}
 \end{aligned} \tag{2}$$

where $x_{t,i,r}$ is the estimated log-conductivity at time step t , node i and realization r , $x_{ref,i}$ is the reference log-conductivity at node i , N_b is the number of nodes, N_e is the number of realizations and $\sigma_{x_{t,i}}^2$ is the variance over the ensemble at time step t and location i . We can see that as assimilation time advances, the $AAB(x)_t$ and the $AESP(x)_t$ decrease and approach a stable value. Clearly, the $AAB(x)_{60}$ corresponding to the NS-EnKF is smaller than that for the standard EnKF, indicating a more accurate estimation. On the contrary, the $AESP(x)_{60}$ corresponding to the NS-EnKF is bigger than that for the standard EnKF, indicating an estimate with higher uncertainty. The discrepancy between $AAB(x)_{60}$ and $AESP(x)_{60}$ is much larger for the standard EnKF than for NS-EnKF, what, according to Chen and Zhang (2006), indicates that the former underestimates the uncertainty in relation with the latter.

Figure 6 displays $\ln K$ histograms of the reference field, the prior ensemble and the updated ensemble by the EnKF and the NS-EnKF after piezometric heads are assimilated. The global mean and standard deviation are preserved during data assimilation for both the standard EnKF and the NS-EnKF. The bimodal histogram for the prior ensemble (related with the sand and shale facies) is correctly preserved by the NS-EnKF during data assimilation, but it is not with the standard EnKF. In the latter case, the updated histogram tends to be Gaussian and the bimodality is almost gone. Besides, the histogram obtained with the standard EnKF shows extremely high and low values, outside of the range of the reference histogram.

Connectivity plays a key role in solute transport simulation and thus a series of measures has been proposed to assess it. Here we measure connectivity by the the probability that two points within the same facies are connected by a continuous path (Stauffer and Aharony, 1994). We focus on the connectivity of sand along the x direction which is the main orientation of the channels. The log-conductivity values are first converted to indicator variables according to

$$I(x) = \begin{cases} 1, & \text{if } \ln K \geq 0 \\ 0, & \text{otherwise,} \end{cases} \tag{3}$$

which, in this particular case, is equivalent to partition them into sand and shale. Then, the program CONNEC3D (Pardo-Igúzquiza and Dowd, 2003) is used to compute the connectivity functions for different

distance lags for log-conductivities larger than $0 \ln(\text{m/d})$. Figure 7 shows connectivity as a function of distance before (Fig.7A) and after data assimilation with the standard EnKF (Fig.7B) and the NS-EnKF (Fig.7C). These figures show the functions for the reference field, for each updated stochastic realization and they also show the mean of the connectivity functions over the stochastic realizations. We can see that the reference connectivity function is underestimated by both the standard EnKF and the NS-EnKF over the considered lags, but the latter performs a little better since the span of the ensemble of realizations includes almost completely the connectivity function for the reference field at long lags.

A random realization in the ensemble and the corresponding connectivity function is shown in Figure 8. In the updated model by the standard EnKF, the border between facies is smoothed (the first graph of the third row), i.e., some values between the two facies (around 0s) are observed, which is reinforced by the histogram in Figure 6C. At the same time, the proportion of being sand is increased (the second graph of the third row), 0.38 in comparison with 0.31 of the reference, indicating more channels while these channels are not connected as well as those in the updated model by the NS-EnKF.

4.2. Prediction capacity of updated $\ln K$

The previous assessments focused on the estimation of log-conductivity itself, for that purpose, the NS-EnKF is found to perform properly in detecting the log-conductivity pattern, preserving the bimodal histogram and estimating the connectivity. Now, the updated log-conductivity realizations will be tested for their ability to perform predictions.

Figure 9 shows the hydraulic head evolution with time at two of the piezometers, one located in shale (Piezometer #19) and the other in sand (Piezometer #44). Hydraulic head predictions for the prior and updated realizations are displayed. Data assimilation by the standard EnKF and the NS-EnKF result in a significant reduction of prediction uncertainty for both methods at both piezometers. The accuracy of hydraulic head prediction is very similar for the NS-EnKF and for the standard EnKF although the updated $\ln K$ fields by NS-EnKF are superior in terms of detection of the spatial patterns of hydraulic conductivities. This fact can be attributed to the smoothing effect of the groundwater flow equation (Delhomme, 1979), i.e., similar piezometric head distributions can be obtained for different $\ln K$ fields, and thus the non-uniqueness of solutions to the inverse problem (Carrera and Neuman, 1986b).

To further evaluate the updated $\ln K$ fields we performed a transport prediction experiment for the case of steady-state flow. Figure 10 shows the set-up of this synthetic experiment. No flow boundary conditions are defined for the northern and southern boundary and prescribed heads along the western (equal to 0 m) and eastern boundary (equal to -10 m). Conservative particles are released along a vertical line at $x = 19.5$

m and two control planes are located at $x = 110$ m and $x = 290$ m, at which the arrival times are recorded. The random walk particle tracking program RW3D (Fernández-García et al., 2005; Li et al., 2011a,b) is used to solve the conservative transport equation. The integrated breakthrough curves (BTCs) at the two control planes are computed and compared with the prior BTCs and the reference BTCs (Fig.11). For both the standard EnKF and the NS-EnKF, the uncertainty of BTC prediction is significantly reduced, indicating the importance to integrate piezometric head data. With standard EnKF, the bias and uncertainty are reduced compared with the prior at both control planes. However, the ensemble median deviates from the reference and the reference is not enclosed in the 90% confidence interval. The travel time is earlier than that in the reference especially for the control plane A, which can be attributed to the wider channels in the updated model than in the reference field as shown in Figure 8. With the NS-EnKF, not only the bias and prediction uncertainty are significantly reduced at both control planes but also the reference is well represented by the ensemble median, and it is contained within the 90% confidence interval, especially for control plane B.

5. Conclusion

A normal-score transformation is introduced into the EnKF, resulting in a new algorithm, the NS-EnKF (Zhou et al., 2011), in which EnKF is applied to model parameters and states following a marginal Gaussian distributions.

The objective of this paper is to investigate the behavior of the NS-EnKF in identifying $\ln K$ patterns for a synthetic non-Gaussian aquifer with a bimodal $\ln K$ distribution by assimilating hydraulic heads in the absence of $\ln K$ measurements. The standard EnKF is also used for the same synthetic set-up in order to compare its performance against the NS-EnKF. The NS-EnKF gives better results than the standard EnKF in the experiment because: (1) the bimodal histogram is well preserved by the NS-EnKF while it is not by the standard EnKF, (2) parameter uncertainty is underestimated by the standard EnKF with respect to the NS-EnKF, (3) channel connectivity along the x direction is underestimated by both methods but the connectivity functions computed on the ensemble of realizations obtained by NS-EnKF are closer to the connectivity function in the synthetic reference field, and (4) the fate of conservative solute is predicted correctly by the updated $\ln K$ fields in NS-EnKF, while the results by standard EnKF exhibit a certain deviation from the reference. In conclusion, transient piezometric head data carry important information which, in conjunction with a knowledge of the prior histogram of the hydraulic conductivity, permits the characterization of a non-Gaussian $\ln K$ distribution even if no hydraulic conductivity data are available.

Acknowledgements. The authors gratefully acknowledge the financial support by the Spanish Ministry of Science and Innovation through project CGL2011-23295. The first author appreciates the financial aid from China Scholarship Council (CSC No. [2007] 3020).

Arulampalam, M. S., Maskell, S., Gordon, N., Clapp, T., 2002. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on signal processing* 50 (2), 174–188.

Bertino, L., Evensen, G., Wackernagel, H., 2003. Sequential data assimilation techniques in oceanography. *International Statistical Review* 71 (2), 223–241.

Burgers, G., Jan van Leeuwen, P., Evensen, G., 1998. Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review* 126 (6), 1719–1724.

Carrera, J., Neuman, S. P., 1986b. Estimation of aquifer parameters under transient and steady state conditions: 2. uniqueness, stability, and solution algorithms. *Water Resources Research*. 22 (2), 211–227.

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

Delhomme, J. P., 1979. Spatial variability and uncertainty in groundwater flow parameters: A geostatistical approach. *Water Resources Research* 15 (2), 269–280.

Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research* 99 (C5), 10143–10162.

Evensen, G., 2007. *Data assimilation: The ensemble Kalman filter*. Springer Verlag, 279pp.

Fernández-García, D., Illangasekare, T., Rajaram, H., 2005. Differences in the scale dependence of dispersivity and retardation factors estimated from forced-gradient and uniform flow tracer tests in three-dimensional physically and chemically heterogeneous porous media. *Water Resources Research* 41 (3), W03012.

Gómez-Hernández, J. J., Journel, A. G., 1993. Joint sequential simulation of Multi-Gaussian fields. In: Soares, A. (Ed.), *Geostatistics Tróia '92*. Vol. 1. Kluwer Academic Publishers, Dordrecht, pp. 85–94.

Gómez-Hernández, J. J., Wen, X. H., 1998. To be or not to be multi-Gaussian? A reflection on stochastic hydrogeology. *Advances in Water Resources* 21 (1), 47–61.

- Gu, Y., Oliver, D. S., 2006. The ensemble Kalman filter for continuous updating of reservoir simulation models. *Journal of Energy Resources Technology* 128, 79–87.
- Harbaugh, A. W., Banta, E. R., Hill, M. C., McDonald, M. G., 2000. MODFLOW-2000, the U.S. geological survey modular ground-water model — user guide to modularization concepts and the ground-water flow process. Tech. Rep. Open-File Report 00-92, U.S. Department of the Interior, U.S. Geological Survey, Reston, Virginia, 121pp.
- Hendricks Franssen, H. J., Kinzelbach, W., 2008. Real-time groundwater flow modeling with the Ensemble Kalman Filter: Joint estimation for states and parameters and the filter inbreeding problem. *Water Resources Research* 44, W09408.
- Hendricks Franssen, H. J., Kinzelbach, W., 2009. Ensemble Kalman filtering versus sequential self-calibration for inverse modelling of dynamic groundwater flow systems. *Journal of Hydrology* 365 (3-4), 261–274.
- Houtekamer, P. L., Mitchell, H. L., 2001. A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review* 129, 123–137.
- Journel, A. G., Deutsch, C. V., 1993. Entropy and spatial disorder. *Mathematical Geology* 25 (3), 329–355.
- Li, L., Zhou, H., Gómez-Hernández, J. J., 2011a. A comparative study of three-dimensional hydraulic conductivity upscaling at the macrodispersion experiment (MADE) site, Columbus air force base, Mississippi (USA). *Journal of Hydrology* 404 (3-4), 278–293.
- Li, L., Zhou, H., Gómez-Hernández, J. J., 2011b. Transport upscaling using multi-rate mass transfer in three-dimensional highly heterogeneous porous media. *Advances in Water Resources* 34 (4), 478–489.
- Moradkhani, H., Sorooshian, S., Gupta, H. V., Houser, P. R., 2005. Dual state-parameter estimation of hydrological models using ensemble Kalman filter. *Advances in Water Resources* (28), 135–147.
- Naevdal, G., Johnsen, L., Aanonsen, S., Vefring, E., Mar. 2005. Reservoir monitoring and continuous model updating using ensemble kalman filter. *SPE Journal* 10 (1).
- Pardo-Igúzquiza, E., Dowd, P. A., 2003. CONNEC3D: a computer program for connectivity analysis of 3D random set models. *Computers & Geosciences* 29, 775–785.
- Schöniger, A., Nowak, W., Hendricks Franssen, H. J., 2011. Parameter estimation by ensemble Kalman filters with transformed data: approach and application to hydraulic tomography. *Water Resources Research*. Under review.

- 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 Science* (5), 495–510.
- Stauffer, D., Aharony, A., 1994. Introduction to percolation theory. Taylor and Francis, London. 181pp.
- Strébel, 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), 1–21.
- Wen, X., Chen, W., 2006. Real-time reservoir model updating using ensemble Kalman filter: The confirming approach. *SPE Journal* 11 (4), 431–442.
- Wen, X., Chen, W., 2007. Some practical issues on real time reservoir updating using ensemble Kalman filter. *SPE Journal* 12 (2), 156–166.
- Zhou, H., Gómez-Hernández, J. J., Hendricks Franssen, H.-J., Li, L., 2011. An approach to handling non-gaussianity of parameters and state variables in ensemble kalman filtering. *Advances in Water Resources* 34 (7), 844–864, DOI: 10.1016/j.advwatres.2011.04.014.
- Zinn, B., Harvey, C., 2003. When good statistical models of aquifer heterogeneity go bad: A comparison of flow, dispersion, and mass transfer in connected and multivariate Gaussian hydraulic conductivity fields. *Water Resources Research* 39 (3), 1051.

Table 1: Parameters defining the multi-Gaussian random function used to generate the log-conductivities within each of the two facies.

Facies	Variogram type	Mean ($\ln(m/d)$)	Variance ($\ln(m/d)^2$)	λ_x^* (m)	λ_y^* (m)	Sill
sand	exponential	3.0	1.0	144	72	1.0
shale	exponential	-2.0	1.0	72	72	0.35

* ranges in the x and y directions

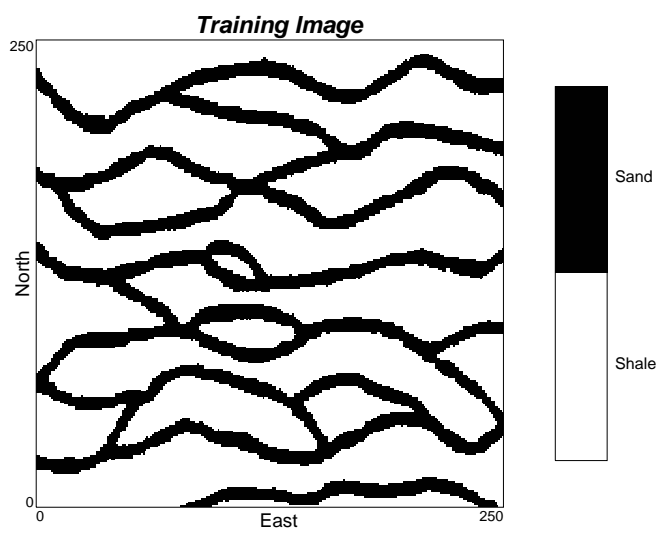


Figure 1: Training image used to generate the binary facies realizations (Strébel, 2000).

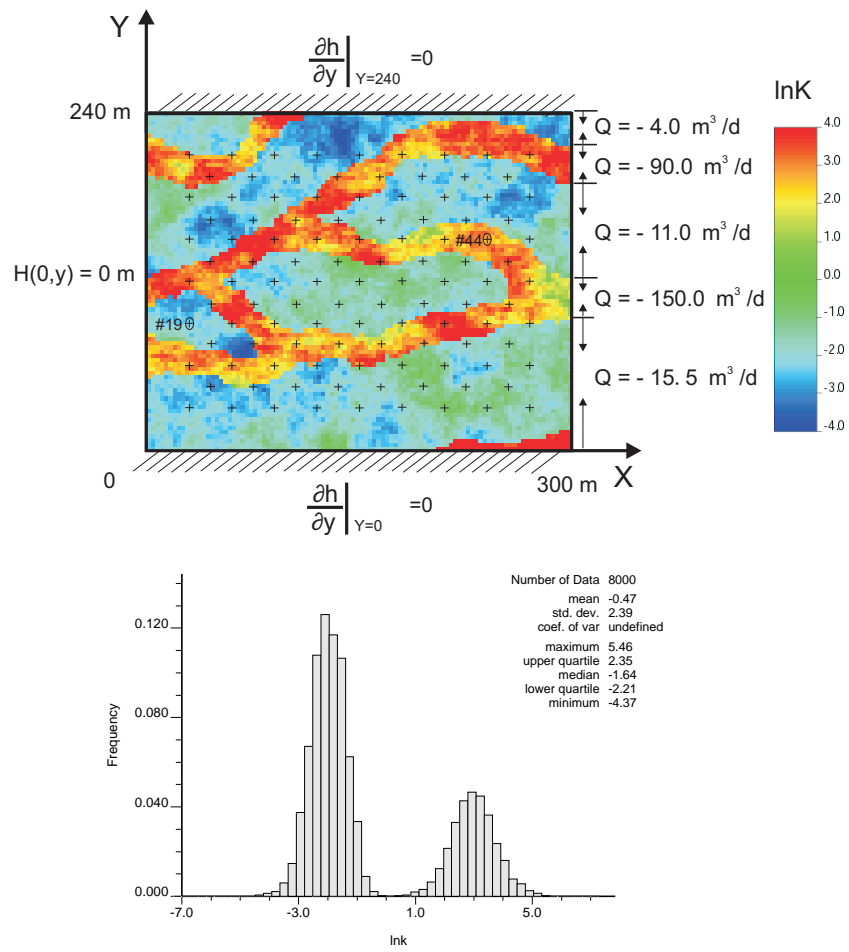


Figure 2: Reference log-conductivity field and histogram. Flow boundary conditions and locations of piezometers are also indicated.

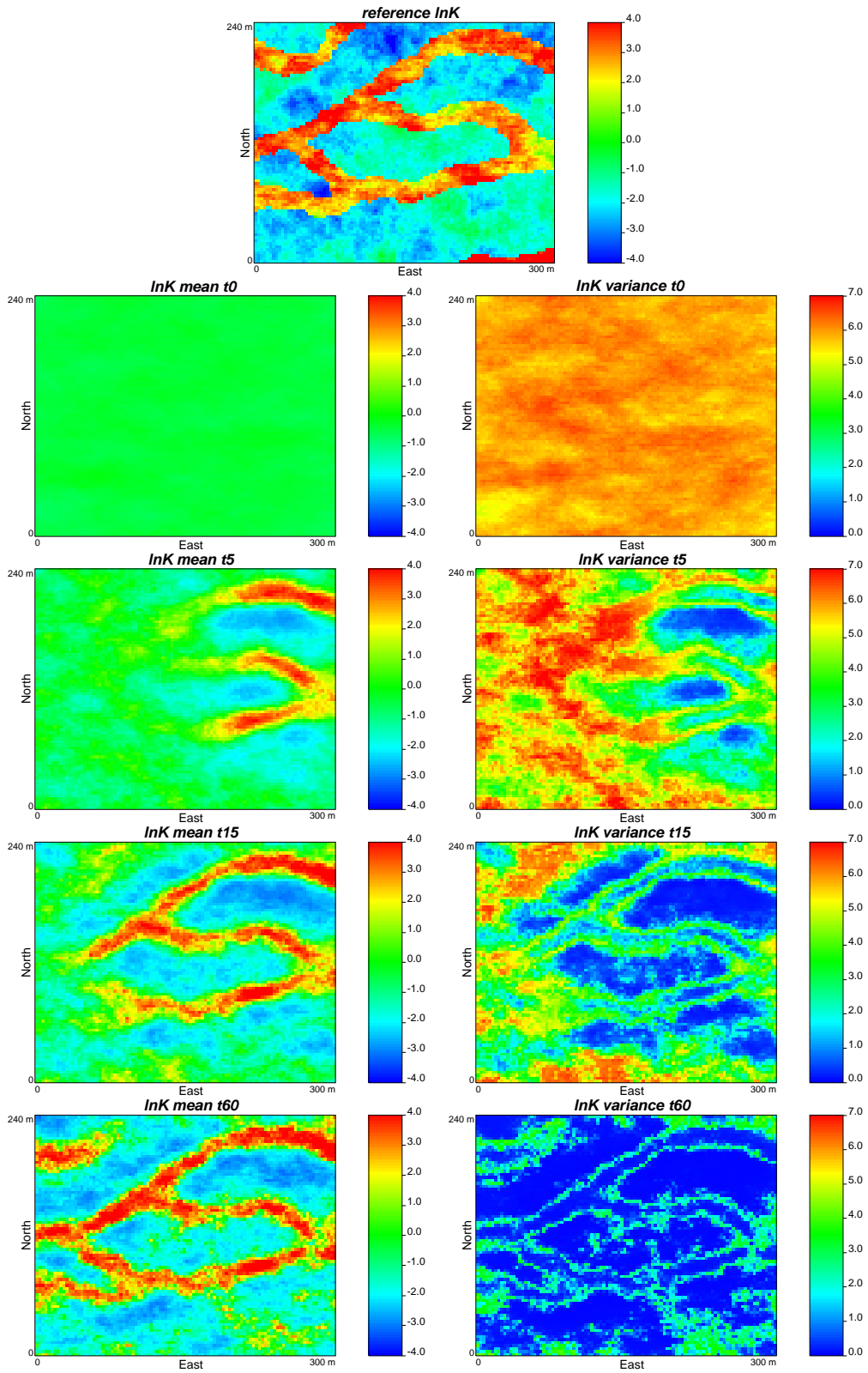


Figure 3: Evolution of the $\ln K$ ensemble mean and variance as data are assimilated with the NS-EnKF for time steps 5, 15 and 60. Reference $\ln K$ is shown for comparison.

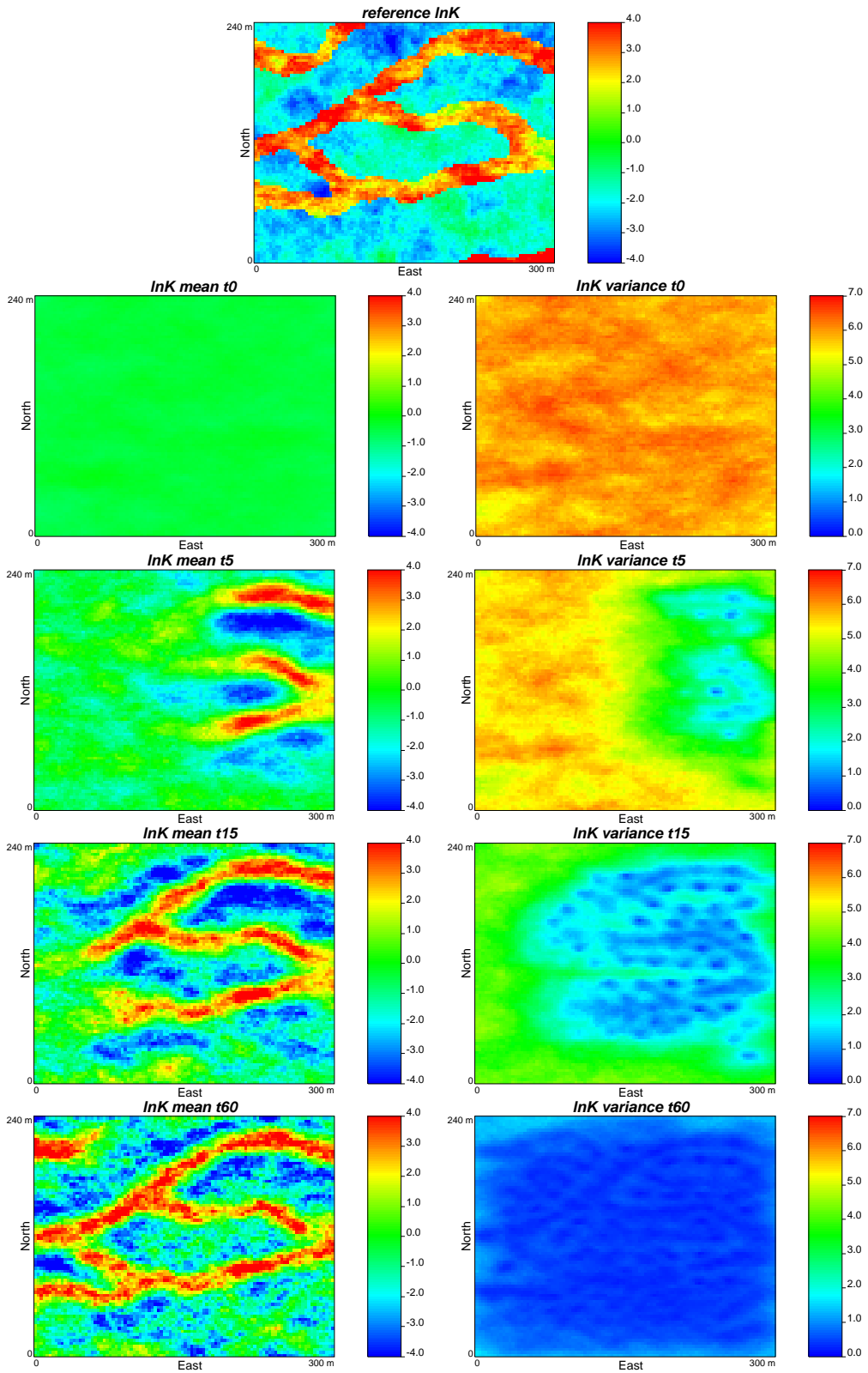


Figure 4: Evolution of the $\ln K$ ensemble mean and variance as data are assimilated with the standard EnKF for time steps 5, 15 and 60. Reference $\ln K$ is shown for comparison.

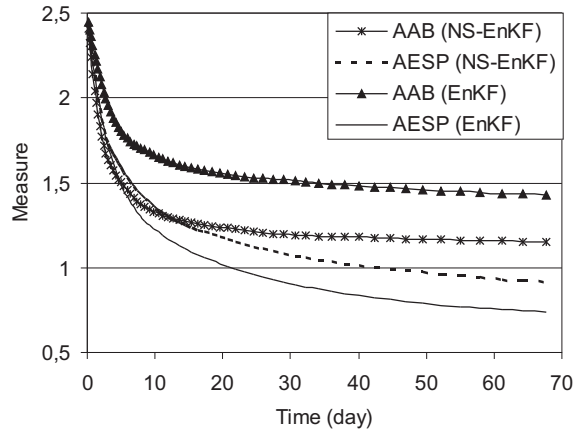


Figure 5: Evolution of Average Absolute Bias AAB(lnK) and Average Ensemble Spread AESP(lnK) for sequential data assimilation with the standard EnKF and the NS-EnKF.

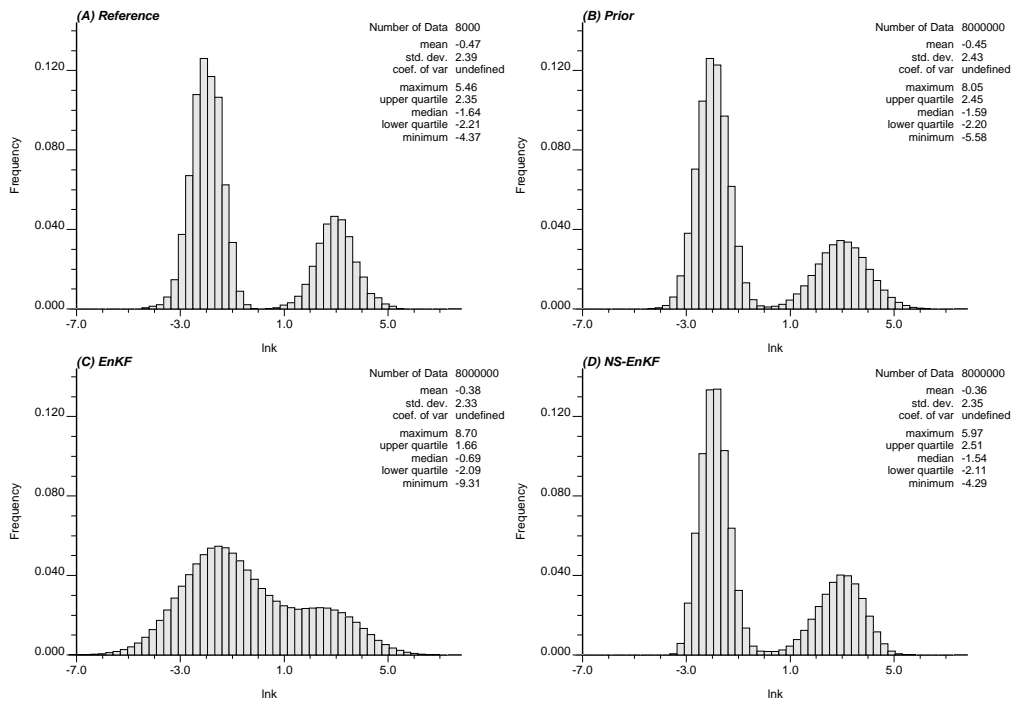


Figure 6: lnK histograms for the (A) reference, (B) prior, (C) posterior with the standard EnKF, and (D) posterior with the NS-EnKF.

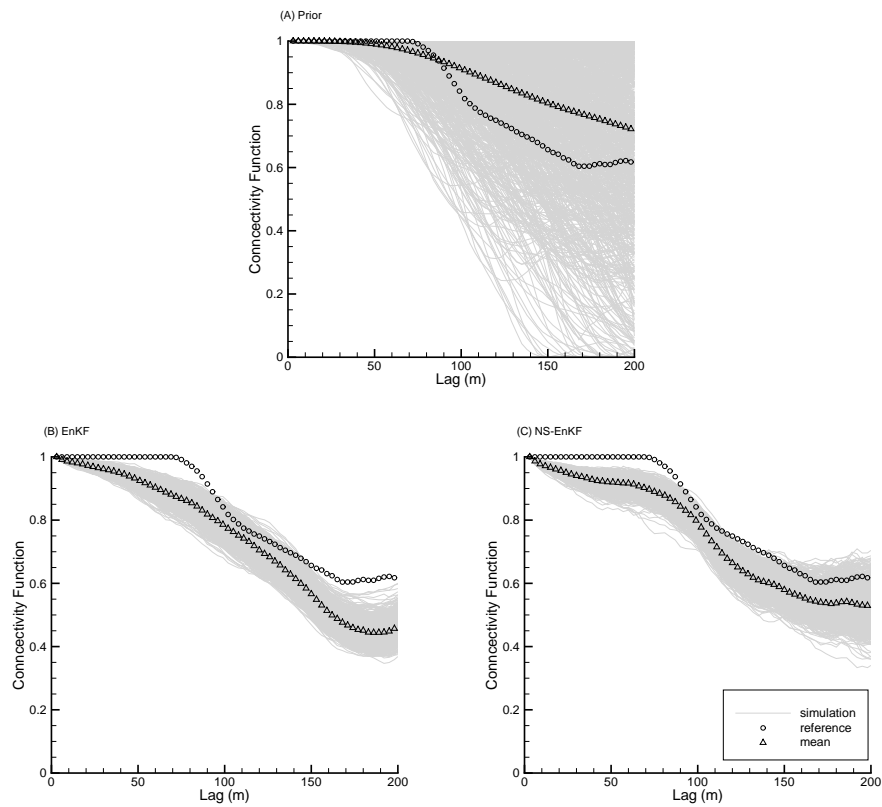


Figure 7: Connectivity as a function of the separation distance, measuring the probability that two points along the x direction are connected by a continuous path of $\ln K$ larger than $0 \ln(m/d)$. Results are shown for (A) the prior unconditional realizations, (B) the realizations conditioned with standard EnKF, and (C) the realizations conditioned with the NS-EnKF.

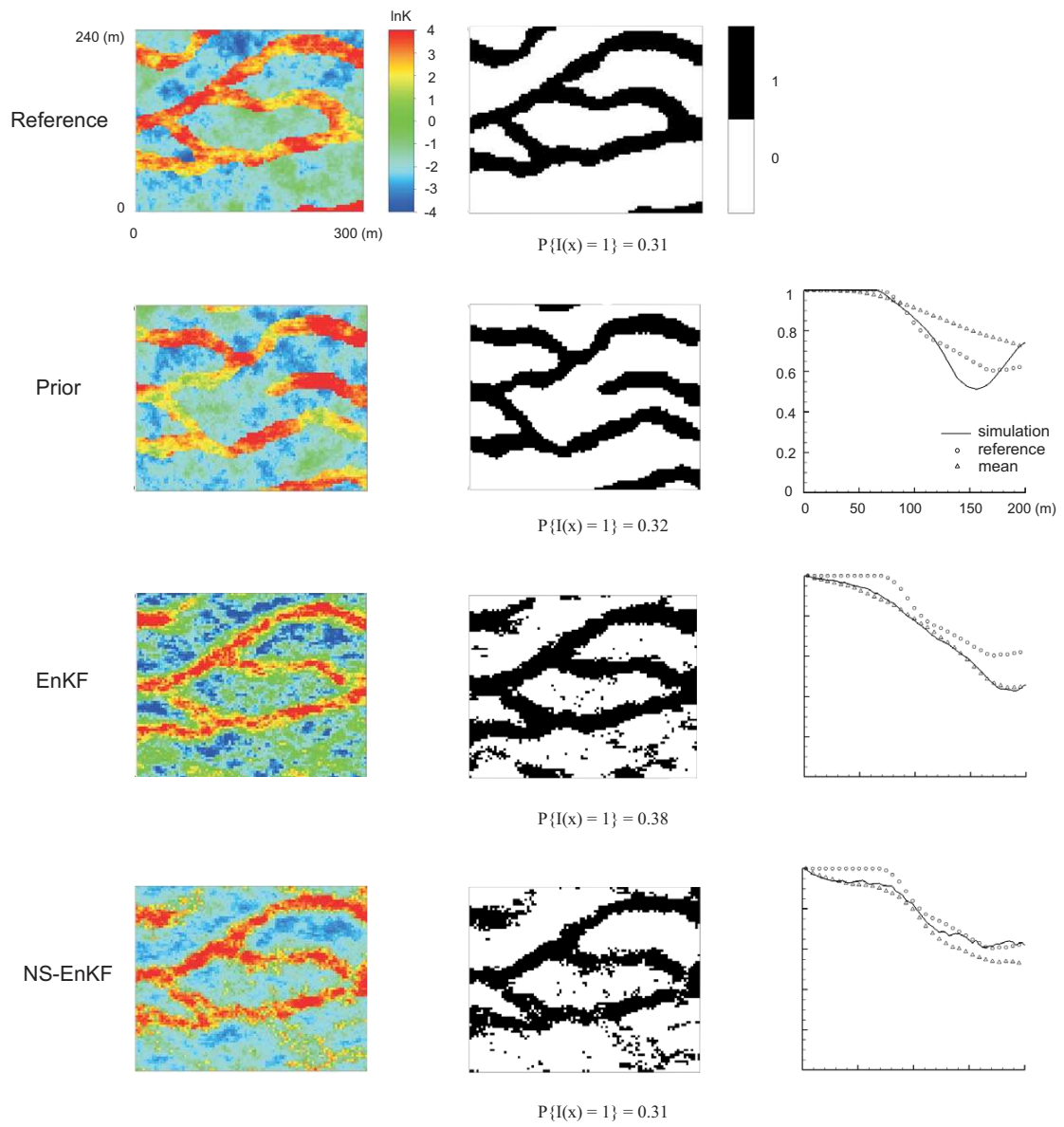


Figure 8: From the top row down: the reference field, initial $\ln K$ distribution for one of the ensemble realizations, final $\ln K$ distribution for the same realization after data assimilation by the EnKF, final $\ln K$ distribution after data assimilation by the NS-EnKF. The first column shows the continuous $\ln K$ values, the second column shows the sand/shale classification given by Equation 3 and the last column shows the connectivity functions for the map in the second column (solid line), for the reference map (top row, second column), and the average of the ensemble of connectivity functions in all realizations. The proportion of sand derived from the maps in the second column is also reported.

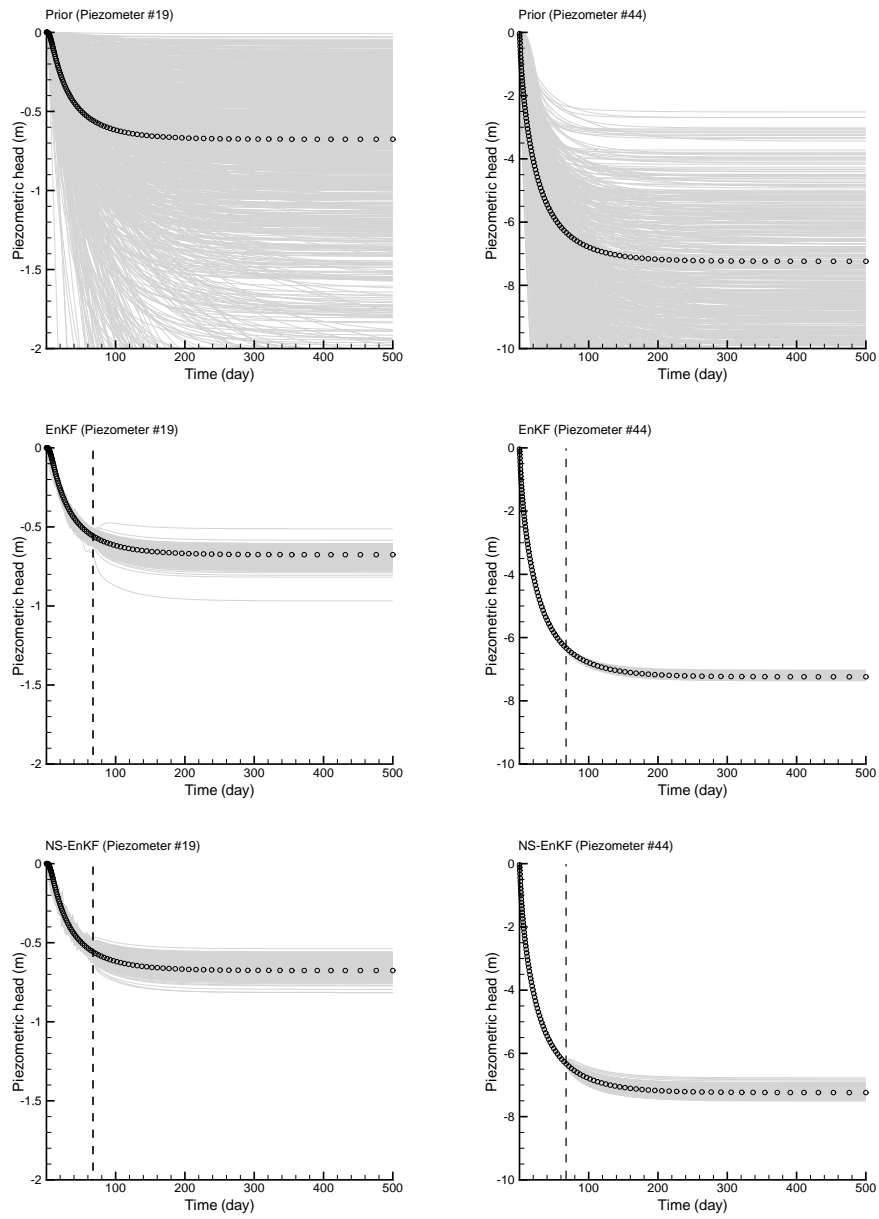


Figure 9: Piezometric head evolution for two of the 111 piezometers, the locations of which are indicated in Figure 2. Results are shown for the prior ensemble (no data assimilation), for assimilation with the standard EnKF and the NS-EnKF (conditioning to piezometric heads until 67.7 days, indicated by the vertical dashed line). Circles refer to piezometric heads in the reference field.

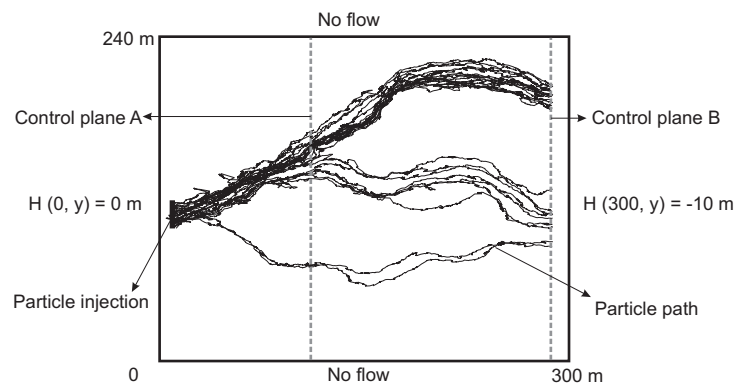


Figure 10: Configuration of the transport prediction experiment. Boundary conditions, locations for particle injection, location of the two control planes and an example of the particle paths in one of the realizations are shown.

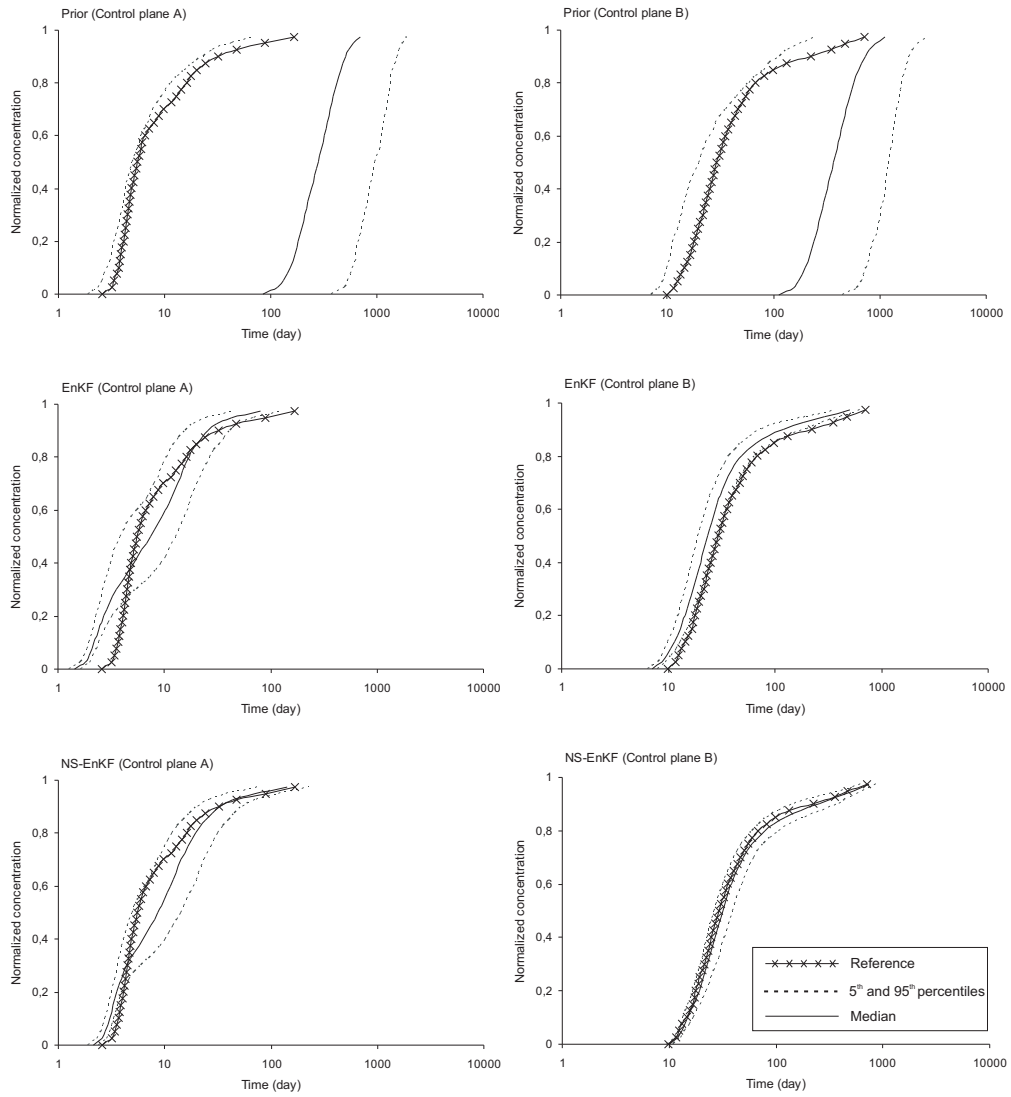


Figure 11: Breakthrough curves at two control planes for the prior ensemble and for the ensemble updated with the standard EnKF and the NS-EnKF. The 5th percentile, median, 95th percentile and reference of the breakthrough curves are shown.